Continuous-Time Signals

Signals occur in a wide range of physical phenomenon. They might be human speech, blood pressure variations with time, seismic waves, radar and sonar signals, pictures or images, stress and strain signals in a building structure, stock market prices, a city's population, or temperature across a plate. These signals are often modeled or represented by a real or complex valued mathematical function of one or more variables. For example, speech is modeled by a function representing air pressure varying with time. The function is acting as a mathematical analogy to the speech signal and, therefore, is called an **analog** signal. For these signals, the independent variable is time and it changes continuously so that the term **continuous-time** signal is also used. In our discussion, we talk of the mathematical function as the signal even though it is really a model or representation of the physical signal.

The description of signals in terms of their sinusoidal frequency content has proven to be one of the most powerful tools of continuous and discrete-time signal description, analysis, and processing. For that reason, we will start the discussion of signals with a development of Fourier transform methods. We will first review the continuous-time methods of the Fourier series (FS), the Fourier transform or integral (FT), and the Laplace transform (LT). Next the discrete-time methods will be developed in more detail with the discrete Fourier transform (DFT) applied to finite length signals followed by the discrete-time Fourier transform (DTFT) for infinitely long signals and ending with the Z-transform which allows the powerful tools of complex variable theory to be applied.

More recently, a new tool has been developed for the analysis of signals. Wavelets and wavelet transforms [link], [link], [link], [link], [link] are another more flexible expansion system that also can describe continuous and discrete-time, finite or infinite duration signals. We will very briefly introduce the ideas behind wavelet-based signal analysis.

The Fourier Series

The problem of expanding a finite length signal in a trigonometric series was posed and studied in the late 1700's by renowned mathematicians such as Bernoulli, d'Alembert, Euler, Lagrange, and Gauss. Indeed, what we now call the Fourier series and the formulas for the coefficients were used by Euler in 1780. However, it was the presentation in 1807 and the paper in 1822 by Fourier stating that an arbitrary function could be represented by a series of sines and cosines that brought the problem to everyone's attention and started serious theoretical investigations and practical applications that continue to this day [link], [link], [link], [link], [link], [link]. The theoretical work has been at the center of analysis and the practical applications have been of major significance in virtually every field of quantitative science and technology. For these reasons and others, the Fourier series is worth our serious attention in a study of signal processing.

Definition of the Fourier Series

We assume that the signal $x(t)$ to be analyzed is well described by a real or complex valued function of a real variable t defined over a finite interval $\{0 \le t \le T\}$. The trigonometric series expansion of $x(t)$ is given by

Equation:

$$x\left(t\right)=\frac{a(0)}{2}+\sum_{k=1}^{\infty}a\left(k\right)\,\cos\left(\frac{2\pi}{T}kt\right)+b\left(k\right)\,\sin\left(\frac{2\pi}{T}kt\right).$$

where $x_k\left(t\right)=cos\left(2\pi kt/T\right)$ and $y_k\left(t\right)=sin\left(2\pi kt/T\right)$ are the basis functions for the expansion. The energy or power in an electrical, mechanical, etc. system is a function of the square of voltage, current, velocity, pressure, etc. For this reason, the natural setting for a representation of signals is the Hilbert space of $L^2\left[0,T\right]$. This modern formulation of the problem is developed in [link], [link]. The sinusoidal basis functions in the trigonometric expansion form a complete orthogonal set in $L^2\left[0,T\right]$. The orthogonality is easily seen from inner products

Equation:

$$\left(\cos\left(\frac{2\pi}{T}kt\right)\,,\,\cos\left(\frac{2\pi}{T}\ell t\right)\right)=\int_0^T\left(\cos\left(\frac{2\pi}{T}kt\right)\,\cos\left(\frac{2\pi}{T}\ell t\right)\right)dt=\delta\left(k-\ell\right)$$

and

Equation:

$$\left(\cos\left(\frac{2\pi}{T}kt\right)\,,\,\sin\left(\frac{2\pi}{T}\ell t\right)\right)=\int_0^T\left(\cos\left(\frac{2\pi}{T}kt\right)\,\sin\left(\frac{2\pi}{T}\ell t\right)\right)dt=0$$

where $\delta(t)$ is the Kronecker delta function with $\delta(0)=1$ and $\delta(k\neq 0)=0$. Because of this, the kth coefficients in the series can be found by taking the inner product of $x(t)$ with the kth basis functions. This gives for the coefficients

Equation:

$$a\left(k\right)=\frac{2}{T}\int_0^T x\left(t\right)\,\cos\left(\frac{2\pi}{T}kt\right)dt$$

and

Equation:

$$b\left(k\right) = \frac{2}{T}\int_0^T x\left(t\right)\,\sin\left(\frac{2\pi}{T}kt\right)dt$$

where T is the time interval of interest or the period of a periodic signal. Because of the orthogonality of the basis functions, a finite Fourier series formed by truncating the infinite series is an optimal least squared error approximation to $x(t)$. If the finite series is defined by

Equation:

$$\widehat{x}\left(t\right) = \frac{a(0)}{2} + \sum_{k=1}^{N} a\left(k\right)\,\cos\left(\frac{2\pi}{T}kt\right) + b\left(k\right)\,\sin\left(\frac{2\pi}{T}kt\right),$$

the squared error is

Equation:

$$\varepsilon = \frac{1}{T}\int_0^T \left|x\left(t\right) - \widehat{x}\left(t\right)\right|^2 dt$$

which is minimized over all $a(k)$ and $b(k)$ by [link] and [link]. This is an extraordinarily important property.

It follows that if $x\left(t\right) \in L^2\left[0, T\right]$, then the series converges to $x(t)$ in the sense that $\varepsilon \to 0$ as $N \to \infty$[link], [link]. The question of point-wise convergence is more difficult. A sufficient condition that is adequate for most application states: If $f(x)$ is bounded, is piece-wise continuous, and has no more than a finite number of maxima over an interval, the Fourier series converges point-wise to $f(x)$ at all points of continuity and to the arithmetic mean at points of discontinuities. If $f(x)$ is continuous, the series converges uniformly at all points [link], [link], [link].

A useful condition [link], [link] states that if $x(t)$ and its derivatives through the qth derivative are defined and have bounded variation, the Fourier coefficients $a(k)$ and $b(k)$ asymptotically drop off at least as fast as $\frac{1}{k^{q+1}}$ as $k \to \infty$. This ties global rates of convergence of the coefficients to local smoothness conditions of the function.

The form of the Fourier series using both sines and cosines makes determination of the peak value or of the location of a particular frequency term difficult. A different form that explicitly gives the peak value of the sinusoid of that frequency and the location or phase shift of that sinusoid is given by

Equation:

$$x\left(t\right)=\frac{d(0)}{2}+\sum_{k=1}^{\infty}d\left(k\right)\,\cos\left(\frac{2\pi}{T}kt+\theta\left(k\right)\right)$$

and, using Euler's relation and the usual electrical engineering notation of $j=\sqrt{-1}$,
Equation:

$$e^{jx}=\cos\left(x\right)+j\,\sin\left(x\right),$$

the complex exponential form is obtained as
Equation:

$$x\left(t\right)=\sum_{k=-\infty}^{\infty}c\left(k\right)e^{j\frac{2\pi}{T}kt}$$

where
Equation:

$$c(k)=a(k)+j\,b(k).$$

The coefficient equation is
Equation:

$$c\left(k\right)=\frac{1}{T}\int_{0}^{T}x\left(t\right)e^{-j\frac{2\pi}{T}kt}dt$$

The coefficients in these three forms are related by
Equation:

$$|d|^2=|c|^2=a^2+b^2$$

and
Equation:

$$\theta=arg\left\{c\right\}=\tan^{-1}\left(\frac{b}{a}\right)$$

It is easier to evaluate a signal in terms of $c(k)$ or $d(k)$ and $\theta(k)$ than in terms of $a(k)$ and $b(k)$. The first two are polar representation of a complex value and the last is rectangular.

The exponential form is easier to work with mathematically.

Although the function to be expanded is defined only over a specific finite region, the series converges to a function that is defined over the real line and is periodic. It is equal to the original function over the region of definition and is a periodic extension outside of the region. Indeed, one could artificially extend the given function at the outset and then the expansion would converge everywhere.

A Geometric View

It can be very helpful to develop a geometric view of the Fourier series where $x(t)$ is considered to be a vector and the basis functions are the coordinate or basis vectors. The coefficients become the projections of $x(t)$ on the coordinates. The ideas of a measure of distance, size, and orthogonality are important and the definition of error is easy to picture. This is done in [link], [link], [link] using Hilbert space methods.

Properties of the Fourier Series

The properties of the Fourier series are important in applying it to signal analysis and to interpreting it. The main properties are given here using the notation that the Fourier series of a real valued function $x(t)$ over $\{0 \le t \le T\}$ is given by $\mathscr{F}\{x(t)\} = c(k)$ and $\tilde{x}\,(t)$ denotes the periodic extensions of $x(t)$.

1. Linear: $\mathscr{F}\{x+y\} = \mathscr{F}\{x\} + \mathscr{F}\{y\}$
   Idea of superposition. Also scalability: $\mathscr{F}\{ax\} = a\mathscr{F}\{x\}$
2. Extensions of $x(t)$: $\tilde{x}\,(t) = \tilde{x}\,(t+T)$
   $\tilde{x}\,(t)$ is periodic.
3. Even and Odd Parts: $x(t) = u(t) + jv(t)$ and
   $C\,(k) = A\,(k) + jB\,(k) = |C\,(k)|\,e^{j\theta(k)}$

u	v	A	B	$\|C\|$	θ
even	0	even	0	even	0
odd	0	0	odd	even	0

0	even	0	even	even	$\pi/2$
0	odd	odd	0	even	$\pi/2$

4. Convolution: If continuous cyclic convolution is defined by
   Equation:

   $$y(t) \;=\; h(t) \circ x(t) \;=\; \int_0^T \tilde{h}(t-\tau)\, \tilde{x}(\tau)\, d\tau$$

   then $\mathscr{F}\{h(t) \circ x(t)\} \;=\; \mathscr{F}\{h(t)\}\, \mathscr{F}\{x(t)\}$
5. Multiplication: If discrete convolution is defined by
   Equation:

   $$e(n) \;=\; d(n) * c(n) \;=\; \sum_{m=-\infty}^{\infty} d(m)\, c(n-m)$$

   then $\mathscr{F}\{h(t)\, x(t)\} \;=\; \mathscr{F}\{h(t)\} * \mathscr{F}\{x(t)\}$
   This property is the inverse of property 4 and vice versa.
6. Parseval: $\frac{1}{T}\int_0^T |x(t)|^2 dt \;=\; \sum_{k=-\infty}^{\infty} |C(k)|^2$
   This property says the energy calculated in the time domain is the same as that calculated in the frequency (or Fourier) domain.
7. Shift: $\mathscr{F}\{\tilde{x}(t-t_0)\} \;=\; C(k)\, e^{-j2\pi t_0 k/T}$
   A shift in the time domain results in a linear phase shift in the frequency domain.
8. Modulate: $\mathscr{F}\left\{x(t)\, e^{j2\pi Kt/T}\right\} \;=\; C(k-K)$
   Modulation in the time domain results in a shift in the frequency domain. This property is the inverse of property 7.
9. Orthogonality of basis functions:
   Equation:

   $$\int_0^T e^{-j2\pi mt/T}\, e^{j2\pi nt/T}\, dt \;=\; T\,\delta(n-m) \;=\; \begin{cases} T & \text{if } n=m \\ 0 & \text{if } n \neq m. \end{cases}$$

   Orthogonality allows the calculation of coefficients using inner products in [link] and [link]. It also allows Parseval's Theorem in property 6. A relaxed version of orthogonality is called “tight frames" and is important in over-specified systems, especially in wavelets.

Examples

- An example of the Fourier series is the expansion of a square wave signal with period 2π. The expansion is
  Equation:

  $$x\left(t\right)\ =\ \frac{4}{\pi}\left[\sin\left(t\right)+\frac{1}{3}\sin\left(3t\right)+\frac{1}{5}\sin\left(5t\right)\cdots\right].$$

  Because $x(t)$ is odd, there are no cosine terms (all $a(k)=0$) and, because of its symmetries, there are no even harmonics (even k terms are zero). The function is well defined and bounded; its derivative is not, therefore, the coefficients drop off as $\frac{1}{k}$.
- A second example is a triangle wave of period 2π. This is a continuous function where the square wave was not. The expansion of the triangle wave is
  Equation:

  $$x\left(t\right)=\frac{4}{\pi}\left[\sin\left(t\right)-\frac{1}{3^2}\sin\left(3t\right)+\frac{1}{5^2}\sin\left(5t\right)+\cdots\right].$$

  Here the coefficients drop off as $\frac{1}{k^2}$ since the function and its first derivative exist and are bounded.

Note the derivative of a triangle wave is a square wave. Examine the series coefficients to see this. There are many books and web sites on the Fourier series that give insight through examples and demos.

Theorems on the Fourier Series

Four of the most important theorems in the theory of Fourier analysis are the inversion theorem, the convolution theorem, the differentiation theorem, and Parseval's theorem [link].

- The inversion theorem is the truth of the transform pair given in [link], [link], and [link]..
- The convolution theorem is property 4.
- The differentiation theorem says that the transform of the derivative of a function is $j\omega$ times the transform of the function.
- Parseval's theorem is given in property 6.

All of these are based on the orthogonality of the basis function of the Fourier series and integral and all require knowledge of the convergence of the sums and integrals. The practical and theoretical use of Fourier analysis is greatly expanded if use is made of distributions or generalized functions (e.g. Dirac delta functions, $\delta(t)$) [link], [link]. Because energy is an important measure of a function in signal processing applications,

the Hilbert space of L^2 functions is a proper setting for the basic theory and a geometric view can be especially useful [link], [link].

The following theorems and results concern the existence and convergence of the Fourier series and the discrete-time Fourier transform [link]. Details, discussions and proofs can be found in the cited references.

- If $f(x)$ has bounded variation in the interval $(-\pi, \pi)$, the Fourier series corresponding to $f(x)$ converges to the value $f(x)$ at any point within the interval, at which the function is continuous; it converges to the value $\frac{1}{2}[f(x+0)+f(x-0)]$ at any such point at which the function is discontinuous. At the points $\pi, -\pi$ it converges to the value $\frac{1}{2}[f(-\pi+0)+f(\pi-0)]$. [link]
- If $f(x)$ is of bounded variation in $(-\pi, \pi)$, the Fourier series converges to $f(x)$, uniformly in any interval (a, b) in which $f(x)$ is continuous, the continuity at a and b being on both sides. [link]
- If $f(x)$ is of bounded variation in $(-\pi, \pi)$, the Fourier series converges to $\frac{1}{2}[f(x+0)+f(x-0)]$, bounded throughout the interval $(-\pi, \pi)$. [link]
- If $f(x)$ is bounded and if it is continuous in its domain at every point, with the exception of a finite number of points at which it may have ordinary discontinuities, and if the domain may be divided into a finite number of parts, such that in any one of them the function is monotone; or, in other words, the function has only a finite number of maxima and minima in its domain, the Fourier series of $f(x)$ converges to $f(x)$ at points of continuity and to $\frac{1}{2}[f(x+0)+f(x-0)]$ at points of discontinuity. [link], [link]
- If $f(x)$ is such that, when the arbitrarily small neighborhoods of a finite number of points in whose neighborhood $|f(x)|$ has no upper bound have been excluded, $f(x)$ becomes a function with bounded variation, then the Fourier series converges to the value $\frac{1}{2}[f(x+0)+f(x-0)]$, at every point in $(-\pi, \pi)$, except the points of infinite discontinuity of the function, provided the improper integral $\int_{-\pi}^{\pi} f(x)dx$ exist, and is absolutely convergent. [link]
- If f is of bounded variation, the Fourier series of f converges at every point x to the value $[f(x+0)+f(x-0)]/2$. If f is, in addition, continuous at every point of an interval $I=(a, b)$, its Fourier series is uniformly convergent in I. [link]
- If $a(k)$ and $b(k)$ are absolutely summable, the Fourier series converges uniformly to $f(x)$ which is continuous. [link]
- If $a(k)$ and $b(k)$ are square summable, the Fourier series converges to $f(x)$ where it is continuous, but not necessarily uniformly. [link]
- Suppose that $f(x)$ is periodic, of period X, is defined and bounded on $[0, X]$ and that at least one of the following four conditions is satisfied: (i) f is piecewise monotonic on $[0, X]$, (ii) f has a finite number of maxima and minima on $[0, X]$ and a finite number of discontinuities on $[0, X]$, (iii) f is of bounded variation on $[0, X]$, (iv) f is piecewise smooth on $[0, X]$: then it will follow that the Fourier series coefficients

may be defined through the defining integral, using proper Riemann integrals, and that the Fourier series converges to $f(x)$ at a.a.x, to $f(x)$ at each point of continuity of f, and to the value $\frac{1}{2}[f(x^-) + f(x^+)]$ at all x. [link]
- For any $1 \le p < \infty$ and any $f \in C^p\left(S^1\right)$, the partial sums
Equation:

$$S_n \; = \; S_n(f) \; = \; \sum_{|k| \le n} \widehat{f}(k) e_k$$

converge to f, uniformly as $n \to \infty$; in fact, $||S_n - f||_\infty$ is bounded by a constant multiple of $n^{-p+1/2}$. [link]

The Fourier series expansion results in transforming a periodic, continuous time function, $\widetilde{x}(t)$, to two discrete indexed frequency functions, $a(k)$ and $b(k)$ that are not periodic.

The Fourier Transform

Many practical problems in signal analysis involve either infinitely long or very long signals where the Fourier series is not appropriate. For these cases, the Fourier transform (FT) and its inverse (IFT) have been developed. This transform has been used with great success in virtually all quantitative areas of science and technology where the concept of frequency is important. While the Fourier series was used before Fourier worked on it, the Fourier transform seems to be his original idea. It can be derived as an extension of the Fourier series by letting the length or period T increase to infinity or the Fourier transform can be independently defined and then the Fourier series shown to be a special case of it. The latter approach is the more general of the two, but the former is more intuitive [link], [link].

Definition of the Fourier Transform

The Fourier transform (FT) of a real-valued (or complex) function of the real-variable t is defined by
Equation:

$$X(\omega) = \int_{-\infty}^{\infty} x(t)\, e^{-j\omega t}\, dt$$

giving a complex valued function of the real variable ω representing frequency. The inverse Fourier transform (IFT) is given by
Equation:

$$x\left(t\right)=\frac{1}{2\pi}\int_{-\infty}^{\infty}X\left(\omega\right)e^{j\omega t}\,d\omega.$$

Because of the infinite limits on both integrals, the question of convergence is important. There are useful practical signals that do not have Fourier transforms if only classical functions are allowed because of problems with convergence. The use of delta functions (distributions) in both the time and frequency domains allows a much larger class of signals to be represented [link].

Properties of the Fourier Transform

The properties of the Fourier transform are somewhat parallel to those of the Fourier series and are important in applying it to signal analysis and interpreting it. The main properties are given here using the notation that the FT of a real valued function $x(t)$ over all time t is given by $\mathscr{F}\{x\}=X(\omega)$.

1. Linear: $\mathscr{F}\{x+y\}=\mathscr{F}\{x\}+\mathscr{F}\{y\}$
2. Even and Oddness: if $x(t)=u(t)+jv(t)$ and $X(\omega)=A(\omega)+jB(\omega)$ then

u	v	A	B	$\lvert X\rvert$	θ
even	0	even	0	even	0
odd	0	0	odd	even	0
0	even	0	even	even	$\pi/2$
0	odd	odd	0	even	$\pi/2$

3. Convolution: If continuous convolution is defined by:
   Equation:

$$y\left(t\right)=h\left(t\right)^{*}x\left(t\right)=\int_{-\infty}^{\infty}h\left(t-\tau\right)x\left(\tau\right)d\tau=\int_{-\infty}^{\infty}h\left(\lambda\right)x\left(t-\lambda\right)d\lambda$$

then $\mathscr{F}\{h(t)^{*}x(t)\}=\mathscr{F}\{h(t)\}\mathscr{F}\{x(t)\}$

4. Multiplication: $\mathscr{F}\left\{h\left(t\right)x\left(t\right)\right\} = \frac{1}{2\pi}\mathscr{F}\left\{h\left(t\right)\right\}^{*}\mathscr{F}\left\{x\left(t\right)\right\}$
5. Parseval: $\int_{-\infty}^{\infty}\left|x\left(t\right)\right|^{2}dt = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left|X\left(\omega\right)\right|^{2}d\omega$
6. Shift: $\mathscr{F}\left\{x\left(t-T\right)\right\} = X\left(\omega\right)e^{-j\omega T}$
7. Modulate: $\mathscr{F}\left\{x\left(t\right)e^{j2\pi Kt}\right\} = X\left(\omega-2\pi K\right)$
8. Derivative: $\mathscr{F}\left\{\frac{dx}{dt}\right\} = j\omega X\left(\omega\right)$
9. Stretch: $\mathscr{F}\left\{x\left(at\right)\right\} = \frac{1}{|a|}X\left(\omega/a\right)$
10. Orthogonality: $\int_{-\infty}^{\infty}e^{-j\omega_1 t}e^{j\omega_2 t} = 2\pi\delta\left(\omega_1-\omega_2\right)$

Examples of the Fourier Transform

Deriving a few basic transforms and using the properties allows a large class of signals to be easily studied. Examples of modulation, sampling, and others will be given.

- If $x(t)=\delta(t)$ then $X(\omega)=1$
- If $x(t)=1$ then $X(\omega)=2\pi\delta(\omega)$
- If $x(t)$ is an infinite sequence of delta functions spaced T apart, $x\left(t\right)=\sum_{n=-\infty}^{\infty}\delta\left(t-nT\right)$, its transform is also an infinite sequence of delta functions of weight $2\pi/T$ spaced $2\pi/T$ apart, $X\left(\omega\right)=2\pi\sum_{k=-\infty}^{\infty}\delta\left(\omega-2\pi k/T\right)$.
- Other interesting and illustrative examples can be found in [link], [link].

Note the Fourier transform takes a function of continuous time into a function of continuous frequency, neither function being periodic. If “distribution" or “delta functions" are allowed, the Fourier transform of a periodic function will be a infinitely long string of delta functions with weights that are the Fourier series coefficients.

The Laplace Transform

The Laplace transform can be thought of as a generalization of the Fourier transform in order to include a larger class of functions, to allow the use of complex variable theory, to solve initial value differential equations, and to give a tool for input-output description of linear systems. Its use in system and signal analysis became popular in the 1950's and remains as the central tool for much of continuous time system theory. The question of convergence becomes still more complicated and depends on complex values of s used in the inverse transform which must be in a “region of convergence" (ROC).

Definition of the Laplace Transform

The definition of the Laplace transform (LT) of a real valued function defined over all positive time t is

Equation:

$$F\left(s\right) = \int_{-\infty}^{\infty} f\left(t\right) e^{-st}\, dt$$

and the inverse transform (ILT) is given by the complex contour integral
Equation:

$$f\left(t\right) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F\left(s\right) e^{st}\, ds$$

where $s = \sigma + j\omega$ is a complex variable and the path of integration for the ILT must be in the region of the s plane where the Laplace transform integral converges. This definition is often called the bilateral Laplace transform to distinguish it from the unilateral transform (ULT) which is defined with zero as the lower limit of the forward transform integral [link]. Unless stated otherwise, we will be using the bilateral transform.

Notice that the Laplace transform becomes the Fourier transform on the imaginary axis, for $s = j\omega$. If the ROC includes the $j\omega$ axis, the Fourier transform exists but if it does not, only the Laplace transform of the function exists.

There is a considerable literature on the Laplace transform and its use in continuous-time system theory. We will develop most of these ideas for the discrete-time system in terms of the z-transform later in this chapter and will only briefly consider only the more important properties here.

The unilateral Laplace transform cannot be used if useful parts of the signal exists for negative time. It does not reduce to the Fourier transform for signals that exist for negative time, but if the negative time part of a signal can be neglected, the unilateral transform will converge for a much larger class of function that the bilateral transform will. It also makes the solution of linear, constant coefficient differential equations with initial conditions much easier.

Properties of the Laplace Transform

Many of the properties of the Laplace transform are similar to those for Fourier transform [link], [link], however, the basis functions for the Laplace transform are not orthogonal. Some of the more important ones are:

1. Linear: $\mathscr{L}\{x + y\} = \mathscr{L}\{x\} + \mathscr{L}\{y\}$

2. Convolution: If $y\left(t\right) = h\left(t\right)*x\left(t\right) = \int h\left(t - \tau\right) x\left(\tau\right) d\tau$
   then $\mathscr{L}\{h(t)*x(t)\} = \mathscr{L}\{h(t)\}\,\mathscr{L}\{x(t)\}$
3. Derivative: $\mathscr{L}\left\{\frac{dx}{dt}\right\} = s\mathscr{L}\left\{x\left(t\right)\right\}$
4. Derivative (ULT): $\mathscr{L}\left\{\frac{dx}{dt}\right\} = s\mathscr{L}\left\{x\left(t\right)\right\} - x\left(0\right)$
5. Integral: $\mathscr{L}\left\{\int x\left(t\right) dt\right\} = \frac{1}{s}\mathscr{L}\left\{x\left(t\right)\right\}$
6. Shift: $\mathscr{L}\left\{x\left(t - T\right)\right\} = C\left(k\right) e^{-Ts}$
7. Modulate: $\mathscr{L}\left\{x\left(t\right) e^{j\omega_0 t}\right\} = X\left(s - j\omega_0\right)$

Examples can be found in [link], [link] and are similar to those of the z-transform presented later in these notes. Indeed, note the parallals and differences in the Fourier series, Fourier transform, and Z-transform.

Discrete-Time Signals

Although the discrete-time signal $x(n)$ could be any ordered sequence of numbers, they are usually samples of a continuous-time signal. In this case, the real or imaginary valued mathematical function $x(n)$ of the integer n is not used as an analogy of a physical signal, but as some representation of it (such as samples). In some cases, the term **digital** signal is used interchangeably with discrete-time signal, or the label digital signal may be use if the function is not real valued but takes values consistent with some hardware system.

Indeed, our very use of the term “discrete-time" indicates the probable origin of the signals when, in fact, the independent variable could be length or any other variable or simply an ordering index. The term “digital" indicates the signal is probably going to be created, processed, or stored using digital hardware. As in the continuous-time case, the Fourier transform will again be our primary tool [link], [link], [link].

Notation has been an important element in mathematics. In some cases, discrete-time signals are best denoted as a sequence of values, in other cases, a vector is created with elements which are the sequence values. In still other cases, a polynomial is formed with the sequence values as coefficients for a complex variable. The vector formulation allows the use of linear algebra and the polynomial formulation allows the use of complex variable theory.

The Discrete Fourier Transform

The description of signals in terms of their sinusoidal frequency content has proven to be as powerful and informative for discrete-time signals as it has for continuous-time signals. It is also probably the most powerful computational tool we will use. We now develop the basic discrete-time methods starting with the discrete Fourier transform (DFT) applied to finite length signals, followed by the discrete-time Fourier transform (DTFT) for infinitely long signals, and ending with the z-transform which uses the powerful tools of complex variable theory.

Definition of the DFT

It is assumed that the signal $x(n)$ to be analyzed is a sequence of N real or complex values which are a function of the integer variable n. The DFT of $x(n)$, also called the spectrum of $x(n)$, is a length N sequence of complex numbers denoted $C(k)$ and defined by

Equation:

$$C\left(k\right)\;=\;\sum_{n=0}^{N-1}x\left(n\right)\,e^{-j\frac{2\pi}{N}nk}$$

using the usual engineering notation: $j=\sqrt{-1}$. The inverse transform (IDFT) which retrieves $x(n)$ from $C(k)$ is given by

Equation:

$$x\left(n\right)\;=\;\frac{1}{N}\sum_{k=0}^{N-1}C\left(k\right)\,e^{j\frac{2\pi}{N}nk}$$

which is easily verified by substitution into [link]. Indeed, this verification will require using the orthogonality of the basis function of the DFT which is

Equation:

$$\sum_{k=0}^{N-1}e^{-j\frac{2\pi}{N}mk}e^{j\frac{2\pi}{N}nk}\;=\;\begin{cases}N & \text{if } n=m\\ 0 & \text{if } n\neq m.\end{cases}$$

The exponential basis functions, $e^{-j\frac{2\pi}{N}k}$, for $k\in\{0,N-1\}$, are the N values of the Nth roots of unity (the N zeros of the polynomial $(s-1)^{N}$). This property is what connects the DFT to convolution and allows efficient algorithms for calculation to be developed [link]. They are used so often that the following notation is defined by

Equation:

$$W_{N}\;=\;e^{-j\frac{2\pi}{N}}$$

with the subscript being omitted if the sequence length is obvious from context. Using this notation, the DFT becomes

Equation:

$$C\left(k\right)\;=\;\sum_{n=0}^{N-1}x\left(n\right)W_{N}^{nk}$$

One should notice that with the finite summation of the DFT, there is no question of convergence or of the ability to interchange the order of summation. No “delta functions” are needed and the N transform values can be calculated exactly (within the

accuracy of the computer or calculator used) from the N signal values with a finite number of arithmetic operations.

Matrix Formulation of the DFT

There are several advantages to using a matrix formulation of the DFT. This is given by writing [link] or [link] in matrix operator form as

Equation:

$$\begin{bmatrix} C_0 \\ C_1 \\ C_2 \\ \vdots \\ C_{N-1} \end{bmatrix} = \begin{bmatrix} W^0 & W^0 & W^0 & \cdots & W^0 \\ W^0 & W^1 & W^2 & & \\ W^0 & W^2 & W^4 & & \\ \vdots & & & & \vdots \\ W^0 & & \cdots & & W^{(N-1)(N-1)} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{bmatrix}$$

or

Equation:

$$\mathbf{C} = \mathbf{F}\mathbf{x}.$$

The orthogonality of the basis function in [link] shows up in this matrix formulation by the columns of F being orthogonal to each other as are the rows. This means that $\mathbf{F}^{\mathbf{T}}\mathbf{F} = k\mathbf{I}$, where k is a scalar constant, and, therefore, $\mathbf{F}^{\mathbf{T}} = k\mathbf{F}^{-\mathbf{1}}$. This is called a unitary operator.

The definition of the DFT in [link] emphasizes the fact that each of the N DFT values are the sum of N products. The matrix formulation in [link] has two interpretations. Each k-th DFT term is the inner product of two vectors, k-th row of $\mathbf{F}$ and $\mathbf{x}$; or, the DFT vector, $\mathbf{C}$ is a weighted sum of the N columns of $\mathbf{F}$ with weights being the elements of the signal vector $\mathbf{x}$. A third view of the DFT is the operator view which is simply the single matrix equation [link].

It is instructive at this point to write a computer program to calculate the DFT of a signal. In Matlab [link], there is a pre-programmed function to calculate the DFT, but that hides the scalar operations. One should program the transform in the scalar interpretive language of Matlab or some other lower level language such as FORTRAN, C, BASIC, Pascal, etc. This will illustrate how many multiplications and

additions and trigonometric evaluations are required and how much memory is needed. Do not use a complex data type which also hides arithmetic, but use Euler's relations

Equation:

$$e^{jx} = \cos(x) + j\sin(x)$$

to explicitly calculate the real and imaginary part of $C(k)$.

If Matlab is available, first program the DFT using only scalar operations. It will require two nested loops and will run rather slowly because the execution of loops is interpreted. Next, program it using vector inner products to calculate each $C(k)$ which will require only one loop and will run faster. Finally, program it using a single matrix multiplication requiring no loops and running much faster. Check the memory requirements of the three approaches.

The DFT and IDFT are a completely well-defined, legitimate transform pair with a sound theoretical basis that do not need to be derived from or interpreted as an approximation to the continuous-time Fourier series or integral. The discrete-time and continuous-time transforms and other tools are related and have parallel properties, but neither depends on the other.

The notation used here is consistent with most of the literature and with the standards given in [link]. The independent index variable n of the signal $x(n)$ is an integer, but it is usually interpreted as time or, occasionally, as distance. The independent index variable k of the DFT $C(k)$ is also an integer, but it is generally considered as frequency. The DFT is called the spectrum of the signal and the magnitude of the complex valued DFT is called the magnitude of that spectrum and the angle or argument is called the phase.

Extensions of x(n)

Although the finite length signal $x(n)$ is defined only over the interval $\{0 \leq n \leq (N-1)\}$, the IDFT of $C(k)$ can be evaluated outside this interval to give well defined values. Indeed, this process gives the periodic property 4. There are two ways of formulating this phenomenon. One is to periodically extend $x(n)$ to $-\infty$ and $+\infty$ and work with this new signal. A second more general way is evaluate all indices n and k modulo N. Rather than considering the periodic extension of $x(n)$ on the line of integers, the finite length line is formed into a circle or a line around a cylinder so that after counting to $N-1$, the next number is zero, not a periodic replication of it. The periodic extension is easier to visualize initially and is more commonly used for

the definition of the DFT, but the evaluation of the indices by residue reduction modulo N is a more general definition and can be better utilized to develop efficient algorithms for calculating the DFT [link].

Since the indices are evaluated only over the basic interval, any values could be assigned $x(n)$ outside that interval. The periodic extension is the choice most consistent with the other properties of the transform, however, it could be assigned to zero [link]. An interesting possibility is to artificially create a length $2N$ sequence by appending $x(-n)$ to the end of $x(n)$. This would remove the discontinuities of periodic extensions of this new length $2N$ signal and perhaps give a more accurate measure of the frequency content of the signal with no artifacts caused by “end effects". Indeed, this modification of the DFT gives what is called the discrete cosine transform (DCT) [link]. We will assume the implicit periodic extensions to $x(n)$ with no special notation unless this characteristic is important, then we will use the notation $\widetilde{x}\,(n)$.

Convolution

Convolution is an important operation in signal processing that is in some ways more complicated in discrete-time signal processing than in continuous-time signal processing and in other ways easier. The basic input-output relation for a discrete-time system is given by so-called linear or non-cyclic convolution defined and denoted by

Equation:

$$y\,(n)\;=\;\sum_{m=-\infty}^{\infty} h\,(m)\,x\,(n-m)\;=\;h\,(n)^{*}x\,(n)$$

where $x(n)$ is the perhaps infinitely long input discrete-time signal, $h(n)$ is the perhaps infinitely long impulse response of the system, and $y(n)$ is the output. The DFT is, however, intimately related to cyclic convolution, not non-cyclic convolution. Cyclic convolution is defined and denoted by

Equation:

$$\tilde{y}\,(n)\;=\;\sum_{m=0}^{N-1} \widetilde{h}\,(m)\,\widetilde{x}\,(n-m)\;=\;h\,(n)\circ x\,(n)$$

where either all of the indices or independent integer variables are evaluated modulo N or all of the signals are periodically extended outside their length N domains.

This cyclic (sometimes called circular) convolution can be expressed as a matrix operation by converting the signal $h(n)$ into a matrix operator as
Equation:

$$\mathbf{H} = \begin{matrix} h_0 & h_{L-1} & h_{L-2} & \cdots & h_1 \\ h_1 & h_0 & h_{L-1} & & \\ h_2 & h_1 & h_0 & & \\ \vdots & & & & \vdots \\ h_{L-1} & & \cdots & & h_0 \end{matrix},$$

The cyclic convolution can then be written in matrix notation as
Equation:

$$\mathbf{Y} = \mathbf{HX}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are column matrices or vectors of the input and output values respectively.

Because non-cyclic convolution is often what you want to do and cyclic convolution is what is related to the powerful DFT, we want to develop a way of doing non-cyclic convolution by doing cyclic convolution.

The convolution of a length N sequence with a length M sequence yields a length $N+M-1$ output sequence. The calculation of non-cyclic convolution by using cyclic convolution requires modifying the signals by appending zeros to them. This will be developed later.

Properties of the DFT

The properties of the DFT are extremely important in applying it to signal analysis and to interpreting it. The main properties are given here using the notation that the DFT of a length-N complex sequence $x(n)$ is $\mathscr{F}\{x(n)\} = C(k)$.

1. Linear Operator: $\mathscr{F}\{x(n)+y(n)\} = \mathscr{F}\{x(n)\} + \mathscr{F}\{y(n)\}$
2. Unitary Operator: $\mathbf{F^{-1}} = \frac{1}{N}\mathbf{F^T}$
3. Periodic Spectrum: $C(k) = C(k+N)$
4. Periodic Extensions of $x(n)$: $x(n) = x(n+N)$

5. Properties of Even and Odd Parts: $x(n) = u(n) + jv(n)$ and $C(k) = A(k) + jB(k)$

u	v	A	B	$\|C\|$	θ
even	0	even	0	even	0
odd	0	0	odd	even	$\pi/2$
0	even	0	even	even	$\pi/2$
0	odd	odd	0	even	0

6. Cyclic Convolution: $\mathscr{F}\{h(n) \circ x(n)\} = \mathscr{F}\{h(n)\}\mathscr{F}\{x(n)\}$
7. Multiplication: $\mathscr{F}\{h(n)x(n)\} = \mathscr{F}\{h(n)\} \circ \mathscr{F}\{x(n)\}$
8. Parseval: $\sum_{n=0}^{N-1} |x(n)|^2 = \frac{1}{N}\sum_{k=0}^{N-1} |C(k)|^2$
9. Shift: $\mathscr{F}\{x(n-M)\} = C(k)e^{-j2\pi Mk/N}$
10. Modulate: $\mathscr{F}\{x(n)e^{j2\pi Kn/N}\} = C(k-K)$
11. Down Sample or Decimate: $\mathscr{F}\{x(Kn)\} = \frac{1}{K}\sum_{m=0}^{K-1} C(k+Lm)$ where $N = LK$
12. Up Sample or Stretch: If $x_s(2n) = x(n)$ for integer n and zero otherwise, then $\mathscr{F}\{x_s(n)\} = C(k)$, for $k = 0, 1, 2, \ldots, 2N-1$
13. N Roots of Unity: $\left(W_N^k\right)^N = 1$ for $k = 0, 1, 2, \ldots, N-1$
14. Orthogonality:
    Equation:

$$\sum_{k=0}^{N-1} e^{-j2\pi mk/N} e^{j2\pi nk/N} = \begin{cases} N & \text{if } n = m \\ 0 & \text{if } n \neq m. \end{cases}$$

15. Diagonalization of Convolution: If cyclic convolution is expressed as a matrix operation by $\mathbf{y} = \mathbf{H}\mathbf{x}$ with $\mathbf{H}$ given by [link], the DFT operator diagonalizes the convolution operator $\mathbf{H}$, or $\mathbf{F}^{\mathbf{T}}\mathbf{H}\mathbf{F} = \mathbf{H_d}$ where $\mathbf{H_d}$ is a diagonal matrix with the N values of the DFT of $h(n)$ on the diagonal. This is a matrix statement of Property 6. Note the columns of $\mathbf{F}$ are the N eigenvectors of $\mathbf{H}$, independent of the values of $h(n)$.

One can show that any “kernel" of a transform that would support cyclic, length-N convolution must be the N roots of unity. This says the DFT is the only transform over the complex number field that will support convolution. However, if one considers various finite fields or rings, an interesting transform, called the Number Theoretic Transform, can be defined and used because the roots of unity are simply two raised to a powers which is a simple word shift for certain binary number representations [link], [link].

Examples of the DFT

It is very important to develop insight and intuition into the DFT or spectral characteristics of various standard signals. A few DFT's of standard signals together with the above properties will give a fairly large set of results. They will also aid in quickly obtaining the DFT of new signals. The discrete-time impulse $\delta(n)$ is defined by

Equation:

$$\delta(n) = \begin{cases} 1 & \text{when } n = 0 \\ 0 & \text{otherwise} \end{cases}$$

The discrete-time pulse $\sqcap_M(n)$ is defined by

Equation:

$$\sqcap_M(n) = \begin{cases} 1 & \text{when } n = 0, 1, \cdots, M-1 \\ 0 & \text{otherwise} \end{cases}$$

Several examples are:

- $DFT\{\delta(n)\} = 1$, The DFT of an impulse is a constant.
- $DFT\{1\} = N\delta(k)$, The DFT of a constant is an impulse.
- $DFT\left\{e^{j2\pi Kn/N}\right\} = N\delta(k-K)$
- $DFT\{\cos(2\pi Mn/N) = \frac{N}{2}[\delta(k-M) + \delta(k+M)]$
- $DFT\{\sqcap_M(n)\} = \frac{\sin\left(\frac{\pi}{N}Mk\right)}{\sin\left(\frac{\pi}{N}k\right)}$

These examples together with the properties can generate a still larger set of interesting and enlightening examples. Matlab can be used to experiment with these results and to gain insight and intuition.

The Discrete-Time Fourier Transform

In addition to finite length signals, there are many practical problems where we must be able to analyze and process essentially infinitely long sequences. For continuous-time signals, the Fourier series is used for finite length signals and the Fourier transform or integral is used for infinitely long signals. For discrete-time signals, we have the DFT for finite length signals and we now present the discrete-time Fourier transform (DTFT) for infinitely long signals or signals that are longer than we want to specify [link]. The DTFT can be developed as an extension of the DFT as N goes to infinity or the DTFT can be independently defined and then the DFT shown to be a special case of it. We will do the latter.

Definition of the DTFT

The DTFT of a possibly infinitely long real (or complex) valued sequence $f(n)$ is defined to be
Equation:

$$F(\omega) \;=\; \sum_{-\infty}^{\infty} f(n)\, e^{-j\omega n}$$

and its inverse denoted IDTFT is given by
Equation:

$$f(n) \;=\; \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\omega)\, e^{j\omega n}\, d\omega.$$

Verification by substitution is more difficult than for the DFT. Here convergence and the interchange of order of the sum and integral are serious questions and have been the topics of research over many years. Discussions of the Fourier transform and series for engineering applications can be found in [link], [link]. It is necessary to allow distributions or delta functions to be used to gain the full benefit of the Fourier transform.

Note that the definition of the DTFT and IDTFT are the same as the definition of the IFS and FS respectively. Since the DTFT is a continuous periodic function of ω, its Fourier series is a discrete set of values which turn out to be the original signal. This duality can be helpful in developing properties and gaining insight into various problems. The conditions on a function to determine if it can be expanded in a FS are

exactly the conditions on a desired frequency response or spectrum that will determine if a signal exists to realize or approximate it.

Properties

The properties of the DTFT are similar to those for the DFT and are important in the analysis and interpretation of long signals. The main properties are given here using the notation that the DTFT of a complex sequence $x(n)$ is $\mathscr{F}\{x(n)\} = X(\omega)$.

1. Linear Operator: $\mathscr{F}\{x+y\} = \mathscr{F}\{x\} + \mathscr{F}\{y\}$
2. Periodic Spectrum: $X(\omega) = X(\omega + 2\pi)$
3. Properties of Even and Odd Parts: $x(n) = u(n) + jv(n)$ and $X(\omega) = A(\omega) + jB(\omega)$

u	v	A	B	$\|X\|$	θ
even	0	even	0	even	0
odd	0	0	odd	even	0
0	even	0	even	even	$\pi/2$
0	odd	odd	0	even	$\pi/2$

4. Convolution: If non-cyclic or linear convolution is defined by: $y(n) = h(n)*x(n) = \sum_{m=-\infty}^{\infty} h(n-m)x(m) = \sum_{k=-\infty}^{\infty} h(k)x(n-k)$ then $\mathscr{F}\{h(n)*x(n)\} = \mathscr{F}\{h(n)\}\mathscr{F}\{x(n)\}$
5. Multiplication: If cyclic convolution is defined by: $Y(\omega) = H(\omega) \circ X(\omega) = \int_0^T \tilde{H}(\omega - \Omega)\tilde{X}(\Omega)d\Omega$ $\mathscr{F}\{h(n)x(n)\} = \frac{1}{2\pi}\mathscr{F}\{h(n)\} \circ \mathscr{F}\{x(n)\}$
6. Parseval: $\sum_{n=-\infty}^{\infty} |x(n)|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |X(\omega)|^2 d\omega$
7. Shift: $\mathscr{F}\{x(n-M)\} = X(\omega)e^{-j\omega M}$
8. Modulate: $\mathscr{F}\left\{x(n)e^{j\omega_0 n}\right\} = X(\omega - \omega_0)$
9. Sample: $\mathscr{F}\{x(Kn)\} = \frac{1}{K}\sum_{m=0}^{K-1} X(\omega + Lm)$ where $N = LK$

10. Stretch: $\mathscr{F}\{x_s(n)\} = X(\omega)$, for $-K\pi \leq \omega \leq K\pi$ where $x_s(Kn) = x(n)$ for integer n and zero otherwise.
11. Orthogonality: $\sum_{n=-\infty}^{\infty} e^{-j\omega_1 n} e^{-j\omega_2 n} = 2\pi\delta(\omega_1 - \omega_2)$

Evaluation of the DTFT by the DFT

If the DTFT of a finite sequence is taken, the result is a continuous function of ω. If the DFT of the same sequence is taken, the results are N evenly spaced samples of the DTFT. In other words, the DTFT of a finite signal can be evaluated at N points with the DFT.

Equation:

$$X(\omega) = DTFT\{x(n)\} = \sum_{n=-\infty}^{\infty} x(n) e^{-j\omega n}$$

and because of the finite length

Equation:

$$X(\omega) = \sum_{n=0}^{N-1} x(n) e^{-j\omega n}.$$

If we evaluate ω at N equally space points, this becomes

Equation:

$$X\left(\frac{2\pi}{N}k\right) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}kn}$$

which is the DFT of $x(n)$. By adding zeros to the end of $x(n)$ and taking a longer DFT, any density of points can be evaluated. This is useful in interpolation and in plotting the spectrum of a finite length signal. This is discussed further in Sampling, Up-Sampling, Down-Sampling, and Multi-Rate Processing.

There is an interesting variation of the Parseval's theorem for the DTFT of a finite length-N signal. If $x(n) \neq 0$ for $0 \geq n \geq N-1$, and if $L \geq N$, then

Equation:

$$\sum_{n=0}^{N-1} |x(n)|^2 = \frac{1}{L}\sum_{k=0}^{L-1} |X(2\pi k/L)|^2 = \frac{1}{\pi}\int_0^{\pi} |X(\omega)|^2 \, d\omega.$$

The second term in [link] says the Riemann sum is equal to its limit in this case.

Examples of DTFT

As was true for the DFT, insight and intuition is developed by understanding the properties and a few examples of the DTFT. Several examples are given below and more can be found in the literature [link], [link], [link]. Remember that while in the case of the DFT signals were defined on the region $\{0 \le n \le (N-1)\}$ and values outside that region were periodic extensions, here the signals are defined over all integers and are not periodic unless explicitly stated. The spectrum is periodic with period 2π.

- $DTFT\{\delta(n)\} = 1$ for all frequencies.
- **Equation:**

$$DTFT\{1\} = 2\pi\delta(\omega)$$

- **Equation:**

$$DTFT\left\{e^{j\omega_0 n}\right\} = 2\pi\delta\left(\omega - \omega_0\right)$$

- **Equation:**

$$DTFT\{\cos(\omega_0 n)\} = \pi\left[\delta\left(\omega - \omega_0\right) + \delta\left(\omega + \omega_0\right)\right]$$

- **Equation:**

$$DTFT\{\sqcap_M(n)\} = \frac{\sin(\omega M k/2)}{\sin(\omega k/2)}$$

The Z-Transform

The z-transform is an extension of the DTFT in a way that is analogous to the Laplace transform for continuous-time signals being an extension of the Fourier transform. It allows the use of complex variable theory and is particularly useful in analyzing and describing systems. The question of convergence becomes still more complicated and

depends on values of z used in the inverse transform which must be in the “region of convergence" (ROC).

Definition of the Z-Transform

The z-transform (ZT) is defined as a polynomial in the complex variable z with the discrete-time signal values as its coefficients [link], [link], [link]. It is given by
Equation:

$$F(z) = \sum_{n=-\infty}^{\infty} f(n) z^{-n}$$

and the inverse transform (IZT) is
Equation:

$$f(n) = \frac{1}{2\pi j} \oint_{ROC} F(z) z^{n-1} dz.$$

The inverse transform can be derived by using the residue theorem [link], [link] from complex variable theory to find $f(0)$ from $z^{-1}F(z)$, $f(1)$ from $F(z)$, $f(2)$ from $zF(z)$, and in general, $f(n)$ from $z^{n-1}F(z)$. Verification by substitution is more difficult than for the DFT or DTFT. Here convergence and the interchange of order of the sum and integral is a serious question that involves values of the complex variable z. The complex contour integral in [link] must be taken in the ROC of the z plane.

A unilateral z-transform is sometimes needed where the definition [link] uses a lower limit on the transform summation of zero. This allow the transformation to converge for some functions where the regular bilateral transform does not, it provides a straightforward way to solve initial condition difference equation problems, and it simplifies the question of finding the ROC. The bilateral z-transform is used more for signal analysis and the unilateral transform is used more for system description and analysis. Unless stated otherwise, we will be using the bilateral z-transform.

Properties

The properties of the ZT are similar to those for the DTFT and DFT and are important in the analysis and interpretation of long signals and in the analysis and description of

discrete-time systems. The main properties are given here using the notation that the ZT of a complex sequence $x(n)$ is $\mathscr{Z}\{x(n)\} = X(z)$.

1. Linear Operator: $\mathscr{Z}\{x+y\} = \mathscr{Z}\{x\} + \mathscr{Z}\{y\}$
2. Relationship of ZT to DTFT: $\mathscr{Z}\{x\}|_{z=e^{j\omega}} = \mathscr{DTFT}\{x\}$
3. Periodic Spectrum: $X\left(e^{j\omega}\right) = X\left(e^{j\omega+2\pi}\right)$
4. Properties of Even and Odd Parts: $x(n) = u(n) + jv(n)$ and $X\left(e^{j\omega}\right) = A\left(e^{j\omega}\right) + jB\left(e^{j\omega}\right)$
   Equation:

$$\begin{array}{cccc} u & v & A & B \\ \\ even & 0 & even & 0 \\ odd & 0 & 0 & odd \\ 0 & even & 0 & even \\ 0 & odd & odd & 0 \end{array}$$

5. Convolution: If discrete non-cyclic convolution is defined by $y(n) = h(n)*x(n) = \sum_{m=-\infty}^{\infty} h(n-m)x(m) = \sum_{k=-\infty}^{\infty} h(k)x(n-k)$ then $\mathscr{Z}\{h(n)*x(n)\} = \mathscr{Z}\{h(n)\}\mathscr{Z}\{x(n)\}$
6. Shift: $\mathscr{Z}\{x(n+M)\} = z^M X(z)$
7. Shift (unilateral): $\mathscr{Z}\{x(n+m)\} = z^m X(z) - z^m x(0) - z^{m-1}x(1) - \cdots - zx(m-1)$
8. Shift (unilateral): $\mathscr{Z}\{x(n-m)\} = z^{-m} X(z) - z^{-m+1} x(-1) - \cdots - x(-m)$
9. Modulate: $\mathscr{Z}\{x(n)a^n\} = X(z/a)$
10. Time mult.: $\mathscr{Z}\{n^m x(n)\} = (-z)^m \frac{d^m X(z)}{dz^m}$
11. Evaluation: The ZT can be evaluated on the unit circle in the z-plane by taking the DTFT of $x(n)$ and if the signal is finite in length, this can be evaluated at sample points by the DFT.

Examples of the Z-Transform

A few examples together with the above properties will enable one to solve and understand a wide variety of problems. These use the unit step function to remove the negative time part of the signal. This function is defined as
Equation:

$$u(n) = \begin{cases} 1 & \text{if } n \geq 0 \\ 0 & \text{if } n < 0 \end{cases}$$

and several bilateral z-transforms are given by

- $\mathscr{Z}\{\delta(n)\} = 1$ for all z.
- $\mathscr{Z}\{u(n)\} = \frac{z}{z-1}$ for $|z| > 1$.
- $\mathscr{Z}\{u(n)a^n\} = \frac{z}{z-a}$ for $|z| > |a|$.

Notice that these are similar to but not the same as a term of a partial fraction expansion.

Inversion of the Z-Transform

The z-transform can be inverted in three ways. The first two have similar procedures with Laplace transformations and the third has no counter part.

- The z-transform can be inverted by the defined contour integral in the ROC of the complex z plane. This integral can be evaluated using the residue theorem [link], [link].
- The z-transform can be inverted by expanding $\frac{1}{z}F(z)$ in a partial fraction expansion followed by use of tables for the first or second order terms.
- The third method is not analytical but numerical. If $F(z) = \frac{P(z)}{Q(z)}$, $f(n)$ can be obtained as the coefficients of long division.

For example
Equation:

$$\frac{z}{z-a} = 1 + a\,z^{-1} + a^2 z^{-2} + \cdots$$

which is $u(n)\,a^n$ as used in the examples above.

We must understand the role of the ROC in the convergence and inversion of the z-transform. We must also see the difference between the one-sided and two-sided transform.

Solution of Difference Equations using the Z-Transform

The z-transform can be used to convert a difference equation into an algebraic equation in the same manner that the Laplace converts a differential equation in to an algebraic equation. The one-sided transform is particularly well suited for solving initial condition problems. The two unilateral shift properties explicitly use the initial values of the unknown variable.

A difference equation DE contains the unknown function $x(n)$ and shifted versions of it such as $x(n-1)$ or $x(n+3)$. The solution of the equation is the determination of $x(t)$. A linear DE has only simple linear combinations of $x(n)$ and its shifts. An example of a linear second order DE is
Equation:

$$a\,x(n) + b\,x(n-1) + c\,x(n-2) \;=\; f(n)$$

A time invariant or index invariant DE requires the coefficients not be a function of n and the linearity requires that they not be a function of $x(n)$. Therefore, the coefficients are constants.

This equation can be analyzed using classical methods completely analogous to those used with differential equations. A solution of the form $x\,(n) = K\lambda^n$ is substituted into the homogeneous difference equation resulting in a second order characteristic equation whose two roots give a solution of the form $x_h\,(n) \;=\; K_1\lambda_1^n + K_2\lambda_2^n$. A particular solution of a form determined by $f(n)$ is found by the method of undetermined coefficients, convolution or some other means. The total solution is the particular solution plus the solution of the homogeneous equation and the three unknown constants K_i are determined from three initial conditions on $x(n)$.

It is possible to solve this difference equation using z-transforms in a similar way to the solving of a differential equation by use of the Laplace transform. The z-transform converts the difference equation into an algebraic equation. Taking the ZT of both sides of the DE gives
Equation:

$$a\,X\,(z) + b\left[z^{-1}X\,(z) + x\,(-1)\right] + c\left[z^{-2}X\,(z) + z^{-1}x\,(-1) + x\,(-2)\right] \;=\; Y\,(z)$$

solving for $X(z)$ gives
Equation:

$$X(z) = \frac{z^2 [Y(z) - b\,x(-1) - x(-2)] - z\,c\,x(-1)}{a\,z^2 + b\,z + c}$$

and inversion of this transform gives the solution $x(n)$. Notice that two initial values were required to give a unique solution just as the classical method needed two values.

These are very general methods. To solve an nth order DE requires only factoring an n th order polynomial and performing a partial fraction expansion, jobs that computers are well suited to. There are problems that crop up if the denominator polynomial has repeated roots or if the transform of $y(n)$ has a root that is the same as the homogeneous equation, but those can be handled with slight modifications giving solutions with terms of the from $n\lambda^n$ just as similar problems gave solutions for differential equations of the form $t\,e^{st}$.

The original DE could be rewritten in a different form by shifting the index to give
Equation:

$$a\,x(n+2) + b\,x(n+1) + c\,x(n) = f(n+2)$$

which can be solved using the second form of the unilateral z-transform shift property.

Region of Convergence for the Z-Transform

Since the inversion integral must be taken in the ROC of the transform, it is necessary to understand how this region is determined and what it means even if the inversion is done by partial fraction expansion or long division. Since all signals created by linear constant coefficient difference equations are sums of geometric sequences (or samples of exponentials), an analysis of these cases will cover most practical situations. Consider a geometric sequence starting at zero.
Equation:

$$f(n) = u(n)\,a^n$$

with a z-transform
Equation:

$$F(z) = 1 + a\,z^{-1} + a^2\,z^{-2} + a^3\,z^{-3} + \cdots + a^M z^{-M}.$$

Multiplying by $a\,z^{-1}$ gives
Equation:

$$a\,z^{-1}F\left(z\right)\;=\;a\,z^{-1}+a^{2}z^{-2}+a^{3}z^{-3}+a^{4}z^{-4}+\cdots+a^{M+1}z^{-M-1}$$

and subtracting from [link] gives
Equation:

$$\left(1-a\,z^{-1}\right)F\left(z\right)\;=\;1-a^{M+1}z^{-M-1}$$

Solving for $F(z)$ results in
Equation:

$$F\left(z\right)\;=\;\frac{1-a^{M+1}z^{-M-1}}{1-a\,z^{-1}}\;=\;\frac{z-a\left(\frac{a}{z}\right)^{M}}{z-a}$$

The limit of this sum as $M\to\infty$ is
Equation:

$$F\left(z\right)\;=\;\frac{z}{z-a}$$

for $|z|>|a|$. This not only establishes the z-transform of $f(n)$ but gives the region in the z plane where the sum converges.

If a similar set of operations is performed on the sequence that exists for negative n
Equation:

$$f\left(n\right)\;=\;u\left(-n-1\right)a^{n}\;=\;\begin{cases}a^{n} & n<0\\ 0 & n\geq 0\end{cases}$$

the result is
Equation:

$$F\left(z\right)\;=\;-\frac{z}{z-a}$$

for $|z| < |a|$. Here we have exactly the same z-transform for a different sequence $f(n)$ but with a different ROC. The pole in $F(z)$ divides the z-plane into two regions that give two different $f(n)$. This is a general result that can be applied to a general rational $F(z)$ with several poles and zeros. The z-plane will be divided into concentric annular regions separated by the poles. The contour integral is evaluated in one of these regions and the poles inside the contour give the part of the solution existing for negative n with the poles outside the contour giving the part of the solution existing for positive n.

Notice that any finite length signal has a z-transform that converges for all z. The ROC is the entire z-plane except perhaps zero and/or infinity.

Relation of the Z-Transform to the DTFT and the DFT

The FS coefficients are weights on the delta functions in a FT of the periodically extended signal. The FT is the LT evaluated on the imaginary axis: $s = j\omega$.

The DFT values are samples of the DTFT of a finite length signal. The DTFT is the z-transform evaluated on the unit circle in the z plane.
Equation:

$$F(z) = \sum_{n=-\infty}^{\infty} x(n) z^{-n} = \mathscr{Z}\mathscr{T}\{x(n)\}$$

Equation:

$$F\left(e^{j\omega}\right) = \sum_{n=-\infty}^{\infty} x(n) e^{-j\omega n} = \mathscr{D}\mathscr{T}\mathscr{F}\mathscr{T}\{x(n)\}$$

and if $x(n)$ is of length N
Equation:

$$F\left(e^{j\frac{2\pi}{N}k}\right) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}kn} = \mathscr{D}\mathscr{F}\mathscr{T}\{x(n)\}$$

It is important to be able to relate the time-domain signal $x(n)$, its spectrum $X(\omega)$, and its z-transform represented by the pole-zero locations on the z plane.

Relationships Among Fourier Transforms

The DFT takes a periodic discrete-time signal into a periodic discrete-frequency representation.

The DTFT takes a discrete-time signal into a periodic continuous-frequency representation.

The FS takes a periodic continuous-time signal into a discrete-frequency representation.

The FT takes a continuous-time signal into a continuous-frequency representation.

The LT takes a continuous-time signal into a function of a continuous complex variable.

The ZT takes a discrete-time signal into a function of a continuous complex variable.

Wavelet-Based Signal Analysis

There are wavelet systems and transforms analogous to the DFT, Fourier series, discrete-time Fourier transform, and the Fourier integral. We will start with the discrete wavelet transform (DWT) which is analogous to the Fourier series and probably should be called the wavelet series [link]. Wavelet analysis can be a form of time-frequency analysis which locates energy or events in time and frequency (or scale) simultaneously. It is somewhat similar to what is called a short-time Fourier transform or a Gabor transform or a windowed Fourier transform.

The history of wavelets and wavelet based signal processing is fairly recent. Its roots in signal expansion go back to early geophysical and image processing methods and in DSP to filter bank theory and subband coding. The current high interest probably started in the late 1980's with the work of Mallat, Daubechies, and others. Since then, the amount of research, publication, and application has exploded. Two excellent descriptions of the history of wavelet research and development are by Hubbard [link] and by Daubechies [link] and a projection into the future by Sweldens [link] and Burrus [link].

The Basic Wavelet Theory

The ideas and foundations of the basic dyadic, multiresolution wavelet systems are now pretty well developed, understood, and available [link], [link], [link], [link]. The **first** basic requirement is that a set of expansion functions (usually a basis) are generated from a single “mother” function by translation and scaling. For the discrete wavelet expansion system, this is

Equation:

$$\varphi_{j,k}\left(t\right)=\varphi\left(2^{j}t-k\right)$$

where j,k are integer indices for the series expansion of the form
Equation:

$$f\left(t\right)=\sum_{j,k}c_{j,k}\,\varphi_{j,k}\left(t\right).$$

The coefficients $c_{j,k}$ are called the discrete wavelet transform of the signal $f(t)$. This use of translation and scale to create an expansion system is the foundation of all so-called first generation wavelets [link].

The system is somewhat similar to the Fourier series described in Equation 51 from Least Squared Error Designed of FIR Filters with frequencies being related by powers of two rather than an integer multiple and the translation by k giving only the two results of cosine and sine for the Fourier series.

The **second** almost universal requirement is that the wavelet system generates a multiresolution analysis (MRA). This means that a low resolution function (low scale j) can be expanded in terms of the same function at a higher resolution (higher j). This is stated by requiring that the generator of a MRA wavelet system, called a scaling function $\varphi(t)$, satisfies
Equation:

$$\varphi\left(t\right)=\sum_{n}h\left(n\right)\varphi\left(2t-n\right).$$

This equation, called the **refinement equation** or the **MRA equation** or **basic recursion equation**, is similar to a differential equation in that its solution is what defines the basic scaling function and wavelet [link], [link].

The current state of the art is that most of the necessary and sufficient conditions on the coefficients $h(n)$ are known for the existence, uniqueness, orthogonality, and other properties of $\varphi(t)$. Some of the theory parallels Fourier theory and some does not.

A **third** important feature of a MRA wavelet system is a discrete wavelet transform (DWT) can be calculated by a digital filter bank using what is now called Mallat's algorithm. Indeed, this connection with digital signal processing (DSP) has been a rich source of ideas and methods. With this filter bank, one can calculate the DWT of a

length-N digital signal with order N operations. This means the number of multiplications and additions grows only linearly with the length of the signal. This compares with $N\log\left(N\right)$ for an FFT and N^2 for most methods and worse than that for some others.

These basic ideas came from the work of Meyer, Daubechies, Mallat, and others but for a time looked like a solution looking for a problem. Then a second phase of research showed there are many problems to which the wavelet is an excellent solution. In particular, the results of Donoho, Johnstone, Coifman, Beylkin, and others opened another set of doors.

Generalization of the Basic Wavelet System

After (in some cases during) much of the development of the above basic ideas, a number of generalizations [link] were made. They are listed below:

1. A larger integer scale factor than $M=2$ can be used to give a more general **M-band** refinement equation [link]
   Equation:
   $$\varphi\left(t\right)=\sum_{n}h\left(n\right)\varphi\left(Mt-n\right)$$
   than the “dyadic" or octave based Equation 4 from Rational Function Approximation. This also gives more than two channels in the accompanying filter bank. It allows a uniform frequency resolution rather than the resulting logarithmic one for $M=2$.
2. The wavelet system called a **wavelet packet** is generated by “iterating" the wavelet branches of the filter bank to give a finer resolution to the wavelet decomposition. This was suggested by Coifman and it too allows a mixture of uniform and logarithmic frequency resolution. It also allows a relatively simple adaptive system to be developed which has an automatically adjustable frequency resolution based on the properties of the signal.
3. The usual requirement of translation orthogonality of the scaling function and wavelets can be relaxed to give what is called a **biorthogonal system**[link]. If the expansion basis is not orthogonal, a dual basis can be created that will allow the usual expansion and coefficient calculations to be made. The main disadvantage is the loss of a Parseval's theorem which maintains energy partitioning. Nevertheless, the greater flexibility of the biorthogonal system allows superior performance in many compression and denoising applications.
4. The basic refinement Equation 4 from Rational Function Approximation gives the scaling function in terms of a compressed version of itself (self-similar). If we

allow two (or more) scaling functions, each being a weighted sum of a compress version of both, a more general set of basis functions results. This can be viewed as a vector of scaling functions with the coefficients being a matrix now. Once again, this generalization allows more flexibility in the characteristics of the individual scaling functions and their related multi-wavelets. These are called **multi-wavelet systems** and are still being developed.
5. One of the very few disadvantages of the discrete wavelet transform is the fact it is not shift invariant. In other words, if you shift a signal in time, its wavelet transform not only shifts, it changes character! For many applications in denoising and compression, this is not desirable although it may be tolerable. The DWT can be made **shift-invariant** by calculating the DWT of a signal for all possible shifts and adding (or averaging) the results. That turns out to be equivalent to removing all of the down-samplers in the associated filter bank (an **undecimated filter bank**), which is also equivalent to building an overdetermined or **redundant DWT** from a traditional wavelet basis. This overcomplete system is similar to a "tight frame" and maintains most of the features of an orthogonal basis yet is shift invariant. It does, however, require $N \log (N)$ operations.
6. Wavelet systems are easily modified to being an adaptive system where the basis adjusts itself to the properties of the signal or the signal class. This is often done by starting with a large collection or library of expansion systems and bases. A subset is adaptively selected based on the efficiency of the representation using a process sometimes called **pursuit**. In other words, a set is chosen that will result in the smallest number of significant expansion coefficients. Clearly, this is signal dependent, which is both its strength and its limitation. It is nonlinear.
7. One of the most powerful structures yet suggested for using wavelets for signal processing is to first take the DWT, then do a point-wise linear or nonlinear processing of the DWT, finally followed by an inverse DWT. Simply setting some of the wavelet domain expansion terms to zero results in linear wavelet domain filtering, similar to what would happen if the same were done with Fourier transforms. Donoho [link], [link] and others have shown by using some form of nonlinear thresholding of the DWT, one can achieve near optimal denoising or compression of a signal. The concentrating or localizing character of the DWT allows this nonlinear thresholding to be very effective.

The present state of activity in wavelet research and application shows great promise based on the above generalizations and extensions of the basic theory and structure [link]. We now have conferences, workshops, articles, newsletters, books, and email groups that are moving the state of the art forward. More details, examples, and software are given in [link], [link], [link].

Discrete-Time Systems

In the context of discussing signal processing, the most general definition of a system is similar to that of a function. A system is a device, formula, rule, or some process that assigns an output signal from some given class to each possible input signal chosen from some allowed class. From this definition one can pose three interesting and practical problems.

1. **Analysis**: If the input signal and the system are given, find the output signal.
2. **Control**: If the system and the output signal are given, find the input signal.
3. **Synthesis**: If the input signal and output signal are given, find the system.

The definition of input and output signal can be quite diverse. They could be scalars, vectors, functions, functionals, or other objects.

All three of these problems are important, but analysis is probably the most basic and its study usually precedes that of the other two. Analysis usually results in a unique solution. Control is often unique but there are some problems where several inputs would give the same output. Synthesis is seldom unique. There are usually many possible systems that will give the same output for a given input.

In order to develop tools for analysis, control, and design of discrete-time systems, specific definitions, restrictions, and classifications must be made. It is the explicit statement of what a system is, not what it isn't, that allows a descriptive theory and design methods to be developed.

Classifications

The basic classifications of signal processing systems are defined and listed here. We will restrict ourselves to discrete-time systems that have ordered sequences of real or complex numbers as inputs and outputs and will denote the input sequence by $x(n)$ and the output sequence by $y(n)$ and show the process of the system by $x(n) \to y(n)$. Although the independent variable n could represent any physical variable, our most common usages causes us to generically call it time but the results obtained certainly are not restricted to this interpretation.

1. **Linear,** A system is classified as linear if two conditions are true.

    - If $x(n) \to y(n)$ then $a\,x(n) \to a\,y(n)$ for all a. This property is called homogeneity or scaling.

- If $x_1(n) \rightarrow y_1(n)$ and $x_2(n) \rightarrow y_2(n)$, then $(x_1(n) + x_2(n)) \rightarrow (y_1(n) + y_2(n))$ for all x_1 and x_2. This property is called superposition or additivity.

   If a system does not satisfy both of these conditions for all inputs, it is classified as nonlinear. For most practical systems, one of these conditions implies the other. Note that a linear system must give a zero output for a zero input.
2. **Time Invariant** , also called index invariant or shift invariant. A system is classified as time invariant if $x(n+k) \rightarrow y(n+k)$ for any integer k. This states that the system responds the same way regardless of when the input is applied. In most cases, the system itself is not a function of time.
3. **Stable** . A system is called bounded-input bounded-output stable if for all bounded inputs, the corresponding outputs are bounded. This means that the output must remain bounded even for inputs artificially constructed to maximize a particular system's output.
4. **Causal** . A system is classified as causal if the output of a system does not precede the input. For linear systems this means that the impulse response of a system is zero for time before the input. This concept implies the interpretation of n as time even though it may not be. A system is semi-causal if after a finite shift in time, the impulse response is zero for negative time. If the impulse response is nonzero for $n \rightarrow -\infty$, the system is absolutely non-causal. Delays are simple to realize in discrete-time systems and semi-causal systems can often be made realizable if a time delay can be tolerated.
5. **Real-Time** . A discrete-time system can operate in “real-time" if an output value in the output sequence can be calculated by the system before the next input arrives. If this is not possible, the input and output must be stored in blocks and the system operates in “batch" mode. In batch mode, each output value can depend on all of the input values and the concept of causality does not apply.

These definitions will allow a powerful class of analysis and design methods to be developed and we start with convolution.

Convolution

The most basic and powerful operation for linear discrete-time system analysis, control, and design is discrete-time convolution. We first define the discrete-time unit impulse, also known as the Kronecker delta function, as

Equation:

$$\delta\left(n\right)=\begin{cases}1 & \text{for }\ n=0\\ 0 & \text{otherwise.}\end{cases}$$

If a system is linear and time-invariant, and $\delta(n)\rightarrow h(n)$, the output $y(n)$ can be calculated from its input $x(n)$ by the operation called convolution denoted and defined by

Equation:

$$y\left(n\right)=h\left(n\right)^{*}x\left(n\right)=\sum_{m=-\infty}^{\infty}h\left(n-m\right)x\left(m\right)$$

It is informative to methodically develop this equation from the basic properties of a linear system.

Derivation of the Convolution Sum

We first define a complete set of orthogonal basis functions by $\delta(n-m)$ for $m=0,1,2,\cdots,\infty$. The input $x(n)$ is broken down into a set of inputs by taking an inner product of the input with each of the basis functions. This produces a set of input components, each of which is a single impulse weighted by a single value of the input sequence $(x(n),\delta(n-m))=x(m)\delta(n-m)$. Using the time invariant property of the system, $\delta(n-m)\rightarrow h(n-m)$ and using the scaling property of a linear system, this gives an output of $x(m)\delta(n-M)\rightarrow x(m)h(n-m)$. We now calculate the output due to $x(n)$ by adding outputs due to each of the resolved inputs using the superposition property of linear systems. This is illustrated by the following diagram:

Equation:

$$x\left(n\right)=\left\{\begin{array}{llll}x(n)\delta(n) & = & x(0)\delta(n) & \rightarrow & x(0)h(n)\\ x(n)\delta(n-1) & = & x(1)\delta(n-1) & \rightarrow & x(1)h(n-1)\\ x(n)\delta(n-2) & = & x(2)\delta(n-2) & \rightarrow & x(2)h(n-2)\\ \vdots & & & & \vdots\\ x(n)\delta(n-m) & = & x(m)\delta(n-m) & \rightarrow & x(m)h(n-m)\end{array}\right\}=y\left(n\right)$$

or
Equation:

$$y\left(n\right)=\sum_{m=-\infty}^{\infty} x\left(m\right)h\left(n-m\right)$$

and changing variables gives
Equation:

$$y\left(n\right)=\sum_{m=-\infty}^{\infty} h\left(n-m\right)x\left(m\right)$$

If the system is linear but time varying, we denote the response to an impulse at $n=m$ by $\delta(n-m)\rightarrow h(n,m)$. In other words, each impulse response may be different depending on when the impulse is applied. From the development above, it is easy to see where the time-invariant property was used and to derive a convolution equation for a time-varying system as
Equation:

$$y\left(n\right)=h\left(n,m\right)*x\left(n\right)=\sum_{m=-\infty}^{\infty} h\left(n,m\right)x\left(m\right).$$

Unfortunately, relaxing the linear constraint destroys the basic structure of the convolution sum and does not result in anything of this form that is useful.

By a change of variables, one can easily show that the convolution sum can also be written
Equation:

$$y\left(n\right)=h\left(n\right)*x\left(n\right)=\sum_{m=-\infty}^{\infty} h\left(m\right)x\left(n-m\right).$$

If the system is causal, $h(n)=0$ for $n<0$ and the upper limit on the summation in Equation 2 from Discrete Time Signals becomes $m=n$. If the input signal is

causal, the lower limit on the summation becomes zero. The form of the convolution sum for a linear, time-invariant, causal discrete-time system with a causal input is
Equation:

$$y(n) = h(n)*x(n) = \sum_{m=0}^{n} h(n-m)x(m)$$

or, showing the operations commute
Equation:

$$y(n) = h(n)*x(n) = \sum_{m=0}^{n} h(m)x(n-m).$$

Convolution is used analytically to analyze linear systems and it can also be used to calculate the output of a system by only knowing its impulse response. This is a very powerful tool because it does not require any detailed knowledge of the system itself. It only uses one experimentally obtainable response. However, this summation cannot only be used to analyze or calculate the response of a given system, it can **be** an implementation of the system. This summation can be implemented in hardware or programmed on a computer and become the signal processor.

The Matrix Formulation of Convolution

Some of the properties and characteristics of convolution and of the systems it represents can be better described by a matrix formulation than by the summation notation. The first L values of the discrete-time convolution defined above can be written as a matrix operator on a vector of inputs to give a vector of the output values.
Equation:

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{L-1} \end{bmatrix} = \begin{bmatrix} h_0 & 0 & 0 & \cdots & 0 \\ h_1 & h_0 & 0 & & \\ h_2 & h_1 & h_0 & & \\ \vdots & & & & \vdots \\ h_{L-1} & & \cdots & & h_0 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{L-1} \end{bmatrix}$$

If the input sequence x is of length N and the operator signal h is of length M, the output is of length $L = N + M - 1$. This is shown for $N = 4$ and $M = 3$ by the rectangular matrix operation

Equation:

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} h_0 & 0 & 0 & 0 \\ h_1 & h_0 & 0 & 0 \\ h_2 & h_1 & h_0 & 0 \\ 0 & h_2 & h_1 & h_0 \\ 0 & 0 & h_2 & h_1 \\ 0 & 0 & 0 & h_2 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

It is clear that if the system is causal ($h(n) = 0$ for $n < 0$), the H matrix is lower triangular. It is also easy to see that the system being time-invariant is equivalent to the matrix being Toeplitz [link]. This formulation makes it obvious that if a certain output were desired from a length 4 input, only 4 of the 6 values could be specified and the other 2 would be controlled by them.

Although the formulation of constructing the matrix from the impulse response of the system and having it operate on the input vector seems most natural, the matrix could have been formulated from the input and the vector would have been the impulse response. Indeed, this might the appropriate formulation if one were specifying the input and output and designing the system.

The basic convolution defined in [link], derived in [link], and given in matrix form in [link] relates the input to the output for linear systems. This is the form of convolution that is related to multiplication of the DTFT and z-transform of signals. However, it is cyclic convolution that is fundamentally related to the DFT and that will be efficiently calculated by the fast Fourier transform (FFT)

developed in Part III of these notes. Matrix formulation of length-L cyclic convolution is given by

Equation:

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{L-1} \end{bmatrix} = \begin{bmatrix} h_0 & h_{L-1} & h_{L-2} & \cdots & h_1 \\ h_1 & h_0 & h_{L-1} & & h_2 \\ h_2 & h_1 & h_0 & & h_3 \\ \vdots & & & & \vdots \\ h_{L-1} & & \cdots & & h_0 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{L-1} \end{bmatrix}$$

This matrix description makes it clear that the matrix operator is always square and the three signals, $x(n)$, $h(n)$, and $y(n)$, are necessarily of the same length.

There are several useful conclusions that can be drawn from linear algebra [link]. The eigenvalues of the non-cyclic are all the same since the eigenvalues of a lower triangular matrix are simply the values on the diagonal.

Although it is less obvious, the eigenvalues of the cyclic convolution matrix are the N values of the DFT of $h(n)$ and the eigenvectors are the basis functions of the DFT which are the column vectors of the DFT matrix. The eigenvectors are completely controlled by the structure of H being a cyclic convolution matrix and are not at all a function of the values of $h(n)$. The DFT matrix equation from [link] is given by

Equation:

$$\mathbf{X} = \mathbf{F}\mathbf{x} \quad \text{and} \quad \mathbf{Y} = \mathbf{F}\mathbf{y}$$

where $\mathbf{X}$ is the length-N vector of the DFT values, $\mathbf{H}$ is the matrix operator for the DFT, and $\mathbf{x}$ is the length-N vector of the signal $x(n)$ values. The same is true for the comparable terms in y.

The matrix form of the length-N cyclic convolution in [link] is written

Equation:

$$\mathbf{y} = \mathbf{H}\mathbf{x}$$

Taking the DFT both sides and using the IDFT on x gives

Equation:

$$\mathbf{Fy} = \mathbf{Y} = \mathbf{FHx} = \mathbf{FHF^{-1}X}$$

If we define the diagonal matrix $\mathbf{H_d}$ as an L by L matrix with the values of the DFT of $h(n)$ on its diagonal, the convolution property of the DFT becomes

Equation:

$$\mathbf{Y} = \mathbf{H_d X}$$

This implies

Equation:

$$\mathbf{H_d} = \mathbf{FHF^{-1}} \quad \text{and} \quad \mathbf{H} = \mathbf{F^{-1}H_dF}$$

which is the basis of the earlier statement that the eigenvalues of the cyclic convolution matrix are the values of the DFT of $h(n)$ and the eigenvectors are the orthogonal columns of $\mathbf{F}$. The DFT matrix diagonalizes the cyclic convolution matrix. This is probably the most concise statement of the relation of the DFT to convolution and to linear systems.

An important practical question is how one calculates the non-cyclic convolution needed by system analysis using the cyclic convolution of the DFT. The answer is easy to see using the matrix description of H. The length of the output of non-cyclic convolution is $N+M-1$. If $N-1$ zeros are appended to the end of $h(n)$ and $M-1$ zeros are appended to the end of $x(n)$, the cyclic convolution of these two augmented signals will produce exactly the same $N+M-1$ values as non-cyclic convolution would. This is illustrated for the example considered before.

Equation:

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} h_0 & 0 & 0 & 0 & h_2 & h_1 \\ h_1 & h_0 & 0 & 0 & 0 & h_2 \\ h_2 & h_1 & h_0 & 0 & 0 & 0 \\ 0 & h_2 & h_1 & h_0 & 0 & 0 \\ 0 & 0 & h_2 & h_1 & h_0 & 0 \\ 0 & 0 & 0 & h_2 & h_1 & h_0 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ 0 \\ 0 \end{bmatrix}$$

Just enough zeros were appended so that the nonzero terms in the upper right-hand corner of $\mathbf{H}$ are multiplied by the zeros in the lower part of $\mathbf{x}$ and, therefore, do not contribute to $\mathbf{y}$. This does require convolving longer signals but the output is exactly what we want and we calculated it with the DFT-compatible cyclic convolution. Note that more zeros could have been appended to h and x and the first $N + M - 1$ terms of the output would have been the same only more calculations would have been necessary. This is sometimes done in order to use forms of the FFT that require that the length be a power of two.

If fewer zeros or none had been appended to h and x, the nonzero terms in the upper right-hand corner of $\mathbf{H}$, which are the “tail" of $h(n)$, would have added the values that would have been at the end of the non-cyclic output of $y(n)$ to the values at the beginning. This is a natural part of cyclic convolution but is destructive if non-cyclic convolution is desired and is called aliasing or folding for obvious reasons. Aliasing is a phenomenon that occurs in several arenas of DSP and the matrix formulation makes it easy to understand.

The Z-Transform Transfer Function

Although the time-domain convolution is the most basic relationship of the input to the output for linear systems, the z-transform is a close second in importance. It gives different insight and a different set of tools for analysis and design of linear time-invariant discrete-time systems.

If our system in linear and time-invariant, we have seen that its output is given by convolution.

Equation:

$$y\left(n\right) = \sum_{m=-\infty}^{\infty} h\left(n-m\right)x\left(m\right)$$

Assuming that $h(n)$ is such that the summation converges properly, we can calculate the output to an input that we already know has a special relation with discrete-time transforms. Let $x\left(n\right) = z^n$ which gives

Equation:

$$y(n) = \sum_{m=-\infty}^{\infty} h(n-m) z^m$$

With the change of variables of $k = n - m$, we have

Equation:

$$y(n) = \sum_{k=-\infty}^{\infty} h(k) z^{n-k} = \left[\sum_{k=-\infty}^{\infty} h(k) z^{-k}\right] z^n$$

or

Equation:

$$y(n) = H(z) z^n$$

We have the remarkable result that for an input of $x(n) = z^n$, we get an output of exactly the same form but multiplied by a constant that depends on z and this constant is the z-transform of the impulse response of the system. In other words, if the system is thought of as a matrix or operator, z^n is analogous to an eigenvector of the system and $H(z)$ is analogous to the corresponding eigenvalue.

We also know from the properties of the z-transform that convolution in the n domain corresponds to multiplication in the z domain. This means that the z-transforms of $x(n)$ and $y(n)$ are related by the simple equation

Equation:

$$Y(z) = H(z) X(z)$$

The z-transform decomposes $x(n)$ into its various components along z^n which passing through the system simply multiplies that value time $H(z)$ and the inverse z-transform recombines the components to give the output. This explains why the z-transform is such a powerful operation in linear discrete-time system theory. Its kernel is the eigenvector of these systems.

The z-transform of the impulse response of a system is called its transfer function (it transfers the input to the output) and multiplying it times the z-transform of the

input gives the z-transform of the output for any system and signal where there is a common region of convergence for the transforms.

Frequency Response of Discrete-Time Systems

The frequency response of a Discrete-Time system is something experimentally measurable and something that is a complete description of a linear, time-invariant system in the same way that the impulse response is. The frequency response of a linear, time-invariant system is defined as the magnitude and phase of the sinusoidal output of the system with a sinusoidal input. More precisely, if

Equation:

$$x(n) = \cos(\omega n)$$

and the output of the system is expressed as

Equation:

$$y(n) = M(\omega)\cos(\omega n + \varphi(\omega)) + T(n)$$

where $T(n)$ contains no components at ω, then $M(\omega)$ is called the magnitude frequency response and $\varphi(\omega)$ is called the phase frequency response. If the system is causal, linear, time-invariant, and stable, $T(n)$ will approach zero as $n \to \infty$ and the only output will be the pure sinusoid at the same frequency as the input. This is because a sinusoid is a special case of z^n and, therefore, an eigenvector.

If z is a complex variable of the special form

Equation:

$$z = e^{j\omega}$$

then using Euler's relation of $e^{jx} = \cos(x) + j\sin(x)$, one has

Equation:

$$x(n) = e^{j\omega n} = \cos(\omega n) + j\sin(\omega n)$$

and therefore, the sinusoidal input of (3.22) is simply the real part of z^n for a particular value of z, and, therefore, the output being sinusoidal is no surprise.

Fundamental Theorem of Linear, Time-Invariant Systems

The fundamental theorem of calculus states that an integral defined as an inverse derivative and one defined as an area under a curve are the same. The fundamental theorem of algebra states that a polynomial given as a sum of weighted powers of the independent variable and as a product of first factors of the zeros are the same. The fundamental theorem of arithmetic states that an integer expressed as a sum of weighted units, tens, hundreds, etc. or as the product of its prime factors is the same.

These fundamental theorems all state equivalences of different ways of expressing or calculating something. The fundamental theorem of linear, time-invariant systems states calculating the output of a system can be done with the impulse response by convolution or with the frequency response (or z-transform) with transforms. Stated another way, it says the frequency response can be found from directly calculating the output from a sinusoidal input or by evaluating the z-transform on the unit circle.

Equation:

$$\mathscr{Z}\left\{h(n)\right\}\big|_{z=e^{j\omega}} \quad = \quad A(\omega)\; e^{j\Theta(\omega)}$$

Pole-Zero Plots

Relation of PZ Plots, FR Plots, Impulse R

State Variable Formulation

Difference Equations

Flow Graph Representation

Standard Structures

FIR and IIR Structures

Quantization Effects

Multidimensional Systems

Sampling, Up--Sampling, Down--Sampling, and Multi--Rate

A very important and fundamental operation in discrete-time signal processing is that of sampling. Discrete-time signals are often obtained from continuous-time signal by simple sampling. This is mathematically modeled as the evaluation of a function of a real variable at discrete values of time [link]. Physically, it is a more complicated and varied process which might be modeled as convolution of the sampled signal by a narrow pulse or an inner product with a basis function or, perhaps, by some nonlinear process.

The sampling of continuous-time signals is reviewed in the recent books by Marks [link] which is a bit casual with mathematical details, but gives a good overview and list of references. He gives a more advanced treatment in [link]. Some of these references are [link], [link], [link], [link], [link], [link], [link]. These will discuss the usual sampling theorem but also interpretations and extensions such as sampling the value and one derivative at each point, or of non uniform sampling.

Multirate discrete-time systems use sampling and sub sampling for a variety of reasons [link], [link]. A very general definition of sampling might be any mapping of a signal into a sequence of numbers. It might be the process of calculating coefficients of an expansion using inner products. A powerful tool is the use of periodically time varying theory, particularly the bifrequency map, block formulation, commutators, filter banks, and multidimensional formulations. One current interest follows from the study of wavelet basis functions. What kind of sampling theory can be developed for signals described in terms of wavelets? Some of the literature can be found in [link], [link], [link], [link], [link].

Another relatively new framework is the idea of tight frames [link], [link], [link]. Here signals are expanded in terms of an over determined set of expansion functions or vectors. If these expansions are what is called a tight frame, the mathematics of calculating the expansion coefficients with inner products works just as if the expansion functions were an orthonormal basis set. The redundancy of tight frames offers interesting possibilities. One example of a tight frame is an over sampled band limited function expansion.

Fourier Techniques

We first start with the most basic sampling ideas based on various forms of Fourier transforms [link], [link], [link].

The Spectrum of a Continuous-Time Signal and the Fourier Transform

Although in many cases digital signal processing views the signal as simple sequence of numbers, here we are going to pose the problem as originating with a function of continuous time. The fundamental tool is the classical Fourier transform defined by

Equation:

$$F\left(\omega\right)=\int f\left(t\right)e^{-j\omega t}\,dt$$

and its inverse

Equation:

$$f\left(t\right)=\frac{1}{2\pi}\int F\left(\omega\right)e^{j\omega t}\,d\omega.$$

where $j=\sqrt{-1}$. The Fourier transform of a signal is called its spectrum and it is complex valued with a magnitude and phase.

If the signal is periodic with period $f(t)=f(t+P)$, the Fourier transform does not exist as a function (it may as a distribution) therefore the spectrum is defined as the set of Fourier series coefficients

Equation:

$$C\left(k\right)=\frac{1}{P}\int_0^P f\left(t\right)e^{-j2\pi kt/P}\,dt$$

with the expansion having the form

Equation:

$$f\left(t\right)=\sum_k C\left(k\right)e^{j2\pi kt/P}.$$

The functions $g_k(t) = e^{j2\pi kt/P}$ form an orthogonal basis for periodic functions and [link] is the inner product $C(k) = \langle f(t), g_k(t) \rangle$.

For the non-periodic case in [link] the spectrum is a function of continuous frequency and for the periodic case in [link], the spectrum is a number sequence (a function of discrete frequency).

The Spectrum of a Sampled Signal and the DTFT

The discrete-time Fourier transform (DTFT) as defined in terms samples of a continuous function is
Equation:

$$F_d(\omega) = \sum_n f(Tn)\, e^{-j\omega Tn}$$

and its inverse
Equation:

$$f(Tn) = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} F_d(\omega)\, e^{j\omega Tn}\, d\omega$$

can be derived by noting that $F_d(\omega)$ is periodic with period $P = 2\pi/T$ and, therefore, it can be expanded in a Fourier series with [link] resulting from calculating the series coefficients using [link].

The spectrum of a discrete-time signal is defined as the DTFT of the samples of a continuous-time signal given in [link]. Samples of the signal are given by the inverse DTFT in [link] but they can also be obtained by directly sampling $f(t)$ in [link] giving
Equation:

$$f(Tn) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)\, e^{j\omega Tn}\, d\omega$$

which can be rewritten as an infinite sum of finite integrals in the form
Equation:

$$f(Tn) = \frac{1}{2\pi} \sum_{\ell} \int_0^{2\pi/T} F(\omega + 2\pi\ell/T)\, e^{j(\omega+2\pi\ell/T)Tn}\, d\omega$$

Equation:

$$= \frac{1}{2\pi} \int_0^{2\pi/T} \left[\sum_{\ell} F(\omega + 2\pi\ell/T) \right] e^{j(\omega+2\pi\ell/T)Tn}\, d\omega$$

where $F_p(\omega)$ is a periodic function made up of shifted versions of $F(\omega)$ (aliased) defined in [link] Because [link] and [link] are equal for all Tn and because the limits can be shifted by π/T without changing the equality, the integrands are equal and we have
Equation:

$$F_d(\omega) = \frac{1}{T} \sum_{\ell} F(\omega + 2\pi\ell/T) = \frac{1}{T} F_p(\omega).$$

where $F_p(\omega)$ is a periodic function made up of shifted versions of $F(\omega)$ as in [link]. The spectrum of the samples of $f(t)$ is an aliased version of the spectrum of $f(t)$ itself. The closer together the samples are taken, the further apart the centers of the aliased spectra are.

This result is very important in determining the frequency domain effects of sampling. It shows what the sampling rate should be and it is the basis for deriving the sampling theorem.

Samples of the Spectrum of a Sampled Signal and the DFT

Samples of the spectrum can be calculated from a finite number of samples of the original continuous-time signal using the DFT. If we let the length of the DFT be N and separation of the samples in the frequency domain be Δ and define the periodic functions

Equation:

$$F_p(\omega) = \sum_{\ell} F(\omega + N\Delta\ell)$$

and
Equation:

$$f_p(t) = \sum_{m} f(t + NTm)$$

then from [link] and [link] samples of the DTFT of $f(Tn)$ are
Equation:

$$F_p(\Delta k) = T\sum_{n} f(Tn)\, e^{-jT\Delta nk}$$

Equation:

$$= T\sum_{m}\sum_{n=0}^{N-1} f(Tn + TNm)\, e^{-j\Delta(Tn+TNm)k}$$

Equation:

$$= T\sum_{n=0}^{N-1}\left[\sum_{m} f(Tn + TNm)\right] e^{-j\Delta(Tn+TNm)k},$$

therefore,
Equation:

$$F_p(\Delta k) \; = \; \mathscr{DFT}\{f_p(Tn)\}$$

if $\Delta TN = 2\pi$. This formula gives a method for approximately calculating values of the Fourier transform of a function by taking the DFT (usually with the FFT) of samples of the function. This formula can easily be verified by forming the Riemann sum to approximate the integrals in [link] or [link].

Samples of the DTFT of a Sequence

If the signal is discrete in origin and is not a sampled function of a continuous variable, the DTFT is defined with $T = 1$ as

Equation:

$$H(\omega) = \sum_n h(n)\, e^{-j\omega n}$$

with an inverse

Equation:

$$h(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H(\omega)\, e^{j\omega n}\, d\omega.$$

If we want to calculate $H(\omega)$, we must sample it and that is written as

Equation:

$$H(\Delta k) = \sum_n h(n)\, e^{-j\Delta kn}$$

which after breaking the sum into an infinite sum of length-N sums as was done in [link] becomes

Equation:

$$H(\Delta k) = \sum_m \sum_{n=0}^{N-1} h(n + Nm)\, e^{-j\Delta kn}$$

if $\Delta = 2\pi/N$. This allows us to calculate samples of the DTFT by taking the DFT of samples of a periodized $h(n)$.

Equation:

$$H(\Delta k) = \mathscr{DFT}\{h_p(n)\}.$$

This a combination of the results in [link] and in [link].

Fourier Series Coefficients from the DFT

If the signal to be analyzed is periodic, the Fourier integral in [link] does not converge to a function (it may to a distribution). This function is usually expanded in a Fourier series to define its spectrum or a frequency description. We will sample this function and show how to approximately calculate the Fourier series coefficients using the DFT of the samples.

Consider a periodic signal $\widetilde{f}(t)=\widetilde{f}(t+P)$ with N samples taken every T seconds to give $\widetilde{Tn}(t)$ for integer n such that $NT=P$. The Fourier series expansion of $\widetilde{f}(t)$ is

Equation:

$$\widetilde{f}(t)=\sum_{k=-\infty}^{\infty}C(k)\,e^{2\pi kt/P}$$

with the coefficients given in [link]. Samples of this are

Equation:

$$\widetilde{f}(Tn)=\sum_{k=-\infty}^{\infty}C(k)\,e^{2\pi kTn/P}=\sum_{k=-\infty}^{\infty}C(k)\,e^{2\pi kn/N}$$

which is broken into a sum of sums as

Equation:

$$\widetilde{f}(Tn)=\sum_{\ell-\infty}^{\infty}\sum_{k=0}^{N-1}C(k+N\ell)\,e^{2\pi(k+N\ell)n/N}=\sum_{k=0}^{N-1}\left[\sum_{\ell-\infty}^{\infty}C(k+N\ell)\right]e^{2\pi kn/N}.$$

But the inverse DFT is of the form

Equation:

$$\widetilde{f}(Tn)=\frac{1}{N}\sum_{k=0}^{N-1}F(k)\,e^{j2\pi nk/N}$$

therefore,
Equation:

$$\mathscr{DFT}\left\{\tilde{f}\left(Tn\right)\right\}=N\sum_{\ell}C\left(k+N\ell\right)=N\,C_p\left(k\right).$$

and we have our result of the relation of the Fourier coefficients to the DFT of a sampled periodic signal. Once again aliasing is a result of sampling.

Shannon's Sampling Theorem

Given a signal modeled as a real (sometimes complex) valued function of a real variable (usually time here), we define a bandlimited function as any function whose Fourier transform or spectrum is zero outside of some finite domain
Equation:

$$|F(\omega)|=0\ \text{ for }\ |\omega|>W$$

for some $W<\infty$. The sampling theorem states that if $f(t)$ is sampled
Equation:

$$f_s\left(n\right)=f\left(Tn\right)$$

such that $T<2\pi/W$, then $f(t)$ can be exactly reconstructed (interpolated) from its samples $f_s\left(n\right)$ using
Equation:

$$f\left(t\right)=\sum_{n=-\infty}^{\infty}f_s\left(n\right)\left[\frac{\sin\left(\pi t/T-\pi n\right)}{\pi t/T-\pi n}\right].$$

This is more compactly written by defining the **sinc** function as
Equation:

$$\text{sinc}\left(x\right)=\frac{\sin\left(x\right)}{x}$$

which gives the sampling formula [Equation 53 from Least Squared Error Design of FIR Filters] the form
Equation:

$$f\left(t\right)=\sum_{n}{f}_{s}\left(n\right)\,\mathrm{sinc}\left(\pi t/T-\pi n\right).$$

The derivation of [Equation 53 from Least Squared Error Design of FIR Filters] or [Equation 56 from Least Squared Error Design of FIR Filters] can be done a number of ways. One of the quickest uses infinite sequences of delta functions and will be developed later in these notes. We will use a more direct method now to better see the assumptions and restrictions.

We first note that if $f(t)$ is bandlimited and if $T<2\pi/W$ then there is no overlap or aliasing in ${F}_{p}\left(\omega\right)$. In other words, we can write [link] as
Equation:

$$f\left(t\right)=\frac{1}{2\pi}\int_{-\infty}^{\infty}F\left(\omega\right){e}^{j\omega t}\,d\omega=\frac{1}{2\pi}\int_{-\pi/T}^{\pi/T}{F}_{p}\left(\omega\right){e}^{j\omega t}\,d\omega$$

but
Equation:

$${F}_{p}\left(\omega\right)=\sum_{\ell}F\left(\omega+2\pi\ell/T\right)=T\sum_{n}f\left(Tn\right){e}^{-j\omega Tn}$$

therefore,
Equation:

$$f\left(t\right)=\frac{1}{2\pi}\int_{-\pi/T}^{\pi/T}\left[T\sum_{n}f\left(Tn\right){e}^{-j\omega Tn}\right]{e}^{j\omega t}\,d\omega$$

Equation:

$$= \frac{T}{2\pi} \sum_n f(Tn) \int_{-\pi/T}^{\pi/T} e^{j(t-Tn)\omega}\, d\omega$$

Equation:

$$= \sum_n f(Tn) \frac{\sin\left(\frac{\pi}{T}t - \pi n\right)}{\frac{\pi}{T}t - \pi n}$$

which is the sampling theorem. An alternate derivation uses a rectangle function and its Fourier transform, the sinc function, together with convolution and multiplication. A still shorter derivation uses strings of delta function with convolutions and multiplications. This is discussed later in these notes.

There are several things to notice about this very important result. First, note that although $f(t)$ is defined for all t from only its samples, it does require an infinite number of them to exactly calculate $f(t)$. Also note that this sum can be thought of as an expansion of $f(t)$ in terms of an orthogonal set of basis function which are the sinc functions. One can show that the coefficients in this expansion of $f(t)$ calculated by an inner product are simply samples of $f(t)$. In other words, the sinc functions span the space of bandlimited functions with a very simple calculation of the expansion coefficients. One can ask the question of what happens if a signal is “under sampled". What happens if the reconstruction formula in Equation 12 from Continuous Time Signals is used when there is aliasing and Equation 57 from Least Squarred Error Design of FIR Filters is not true. We will not pursue that just now. In any case, there are many variations and generalizations of this result that are quite interesting and useful.

Calculation of the Fourier Transform and Fourier Series using the FFT

Most theoretical and mathematical analysis of signals and systems use the Fourier series, Fourier transform, Laplace transform, discrete-time Fourier transform (DTFT), or the z-transform, however, when we want to actually evaluate transforms, we calculate values at sample frequencies. In other words, we use the discrete Fourier transform (DFT) and, for efficiency, usually evaluate it with the FFT algorithm. An important question is how can we calculate or approximately calculate these symbolic formula-based transforms with our

practical finite numerical tool. It would certainly seem that if we wanted the Fourier transform of a signal or function, we could sample the function, take its DFT with the FFT, and have some approximation to samples of the desired Fourier transform. We saw in the previous section that it is, in fact, possible provided some care is taken.

Summary

For the signal that is a function of a continuous variable we have

Equation:

$$\begin{array}{llll} \text{FT:} & f(t) & \to & F(\omega) \\ \text{DTFT:} & f(Tn) & \to & \frac{1}{T}\,F_p\big(\omega\big) = \frac{1}{T}\sum_\ell F\big(\omega + 2\pi\ell/T\big) \\ \text{DFT:} & f_p(Tn) & \to & \frac{1}{T}\,F_p\big(\Delta k\big) \ \text{ for } \ \Delta TN = 2\pi \end{array}$$

For the signal that is a function of a discrete variable we have

Equation:

$$\begin{array}{llll} \text{DTFT:} & h(n) & \to & H(\omega) \\ \text{DFT:} & h_p(n) & \to & H(\Delta k)\text{for}\Delta N = 2\pi \end{array}$$

For the periodic signal of a continuous variable we have

Equation:

$$\begin{array}{llll} \text{FS:} & \tilde{g}(t) & \to & C(k) \\ \text{DFT:} & \tilde{g}(Tn) & \to & N\,C_p(k) \ \text{ for } \ TN = P \end{array}$$

For the sampled bandlimited signal we have

Equation:

$$\begin{array}{llll} \text{Sinc:} & f(t) & \to & f(Tn) \\ & & & f(t) = \sum_n f(Tn)\,\text{sinc}(2\pi t/T - \pi n) \\ & & & \text{if } \ F(\omega) = 0 \ \text{ for } \ |\omega| > 2\pi/T \end{array}$$

These formulas summarize much of the relations of the Fourier transforms of sampled signals and how they might be approximately calculate with the FFT. We next turn to the use of distributions and strings of delta functions as tool to study sampling.

Sampling Functions — the Shah Function

Th preceding discussions used traditional Fourier techniques to develop sampling tools. If distributions or delta functions are allowed, the Fourier transform will exist for a much larger class of signals. One should take care when using distributions as if they were functions but it is a very powerful extension.

There are several functions which have equally spaced sequences of impulses that can be used as tools in deriving a sampling formula. These are called “pitch fork" functions, picket fence functions, comb functions and shah functions. We start first with a finite length sequence to be used with the DFT. We define

Equation:

$$\text{Ш}_M(n) = \sum_{m=0}^{L-1} \delta(n - Mm)$$

where $N = LM$.

Equation:

$$DFT\{\text{Ш}_M(n)\} = \sum_{n=0}^{N-1} \left[\sum_{m=0}^{L-1} \delta(n - Mm)\right] e^{-j2\pi nk/N}$$

Equation:

$$= \sum_{m=0}^{L-1} \left[\sum_{n=0}^{N-1} \delta(n - Mm) e^{-j2\pi nk/N}\right]$$

Equation:

$$= \sum_{m=0}^{L-1} e^{-j2\pi Mmk/N} = \sum_{m=0}^{L-1} e^{-j2\pi mk/L}$$

Equation:

$$= \begin{cases} L & < k > L = 0 \\ 0 & \text{otherwise} \end{cases}$$

Equation:

$$= L \sum_{l=0}^{M-1} \delta\left(k - Ll\right) = L\text{Ш}_L\left(k\right)$$

For the DTFT we have a similar derivation:

Equation:

$$DTFT\left\{\text{Ш}_M\left(n\right)\right\} = \sum_{n=-\infty}^{\infty} \left[\sum_{m=0}^{L-1} \delta\left(n - Mm\right)\right] e^{-j\omega n}$$

Equation:

$$= \sum_{m=0}^{L-1} \left[\sum_{n=-\infty}^{\infty} \delta\left(n - Mm\right) e^{-j\omega n}\right]$$

Equation:

$$= \sum_{m=0}^{L-1} e^{-j\omega Mm}$$

Equation:

$$= \begin{cases} L & \omega = k2\pi/M \\ 0 & \text{otherwise} \end{cases}$$

Equation:

$$= \sum_{l=0}^{M-1} \delta\left(\omega - 2\pi l/Ml\right) = K\,\text{Ш}_{2\pi/\mathrm{M}}\left(\omega\right)$$

where K is constant.

An alternate derivation for the DTFT uses the inverse DTFT.
Equation:

$$IDTFT\left\{\text{Ш}_{2\pi/M}\left(\omega\right)\right\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \text{Ш}_{2\pi/M}\left(\omega\right) e^{j\omega n}\,d\omega$$

Equation:

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{l}\delta\left(\omega - 2\pi l/M\right)e^{j\omega n}\,d\omega$$

Equation:

$$= \frac{1}{2\pi}\sum_{l}\int_{-\pi}^{\pi}\delta\left(\omega - 2\pi l/M\right)e^{j\omega n}\,d\omega$$

Equation:

$$= \frac{1}{2\pi}\sum_{l=0}^{M-1} e^{2\pi l n/M} = \begin{cases} M/2\pi & n = M \\ 0 & \text{otherwise} \end{cases}$$

Equation:

$$= \left(\frac{M}{2\pi}\right)\text{Ш}_{2\pi/\mathrm{M}}\left(\omega\right)$$

Therefore,
Equation:

$$\text{Ш}_{M}\left(n\right) \rightarrow \left(\frac{2\pi}{M}\right)\text{Ш}_{2\pi/\mathrm{T}}\left(\omega\right)$$

For regular Fourier transform, we have a string of impulse functions in both the time and frequency. This we see from:
Equation:

$$FT\left\{\text{Ш}_T(t)\right\} = \int_{-\infty}^{\infty} \sum_n \delta(t - nT)\, e^{-j\omega t}\, dt = \sum_n \int \delta(t - nT)\, e^{-j\omega t}\, dt$$

Equation:

$$= \sum_n e^{-j\omega nT} = \begin{cases} \infty & \omega = 2\pi/T \\ 0 & \text{otherwise} \end{cases}$$

Equation:

$$= \frac{2\pi}{T}\, \text{Ш}_{2\pi/\text{T}}\left(\omega\right)$$

The multiplicative constant is found from knowing the result for a single delta function.

These “shah functions" will be useful in sampling signals in both the continuous time and discrete time cases.

Up–Sampling, Signal Stretching, and Interpolation

In several situations we would like to increase the data rate of a signal or, to increase its length if it has finite length. This may be part of a multi rate system or part of an interpolation process. Consider the process of inserting $M-1$ zeros between each sample of a discrete time signal.
Equation:

$$y\left(n\right) = \begin{cases} x(n/M) & < n > M = 0(\text{or } \ n \ \ = \text{kM}) \\ 0 & \text{otherwise} \end{cases}$$

For the finite length sequence case we calculate the DFT of the stretched or up–sampled sequence by
Equation:

$$C_s(k) = \sum_{n=0}^{MN-1} y(n) W_{MN}^{nk}$$

Equation:

$$C_s(k) = \sum_{n=0}^{MN-1} x(n/M)\ \text{Ш}_M(n) W_{MN}^{nk}$$

where the length is now NM and $k = 0, 1, \cdots, NM - 1$. Changing the index variable $n = Mm$ gives:

Equation:

$$C_s(k) = \sum_{m=0}^{N-1} x(m) W_N^{mk} = C(k).$$

which says the DFT of the stretched sequence is exactly the same as the DFT of the original sequence but over M periods, each of length N.

For up–sampling an infinitely long sequence, we calculate the DTFT of the modified sequence in Equation 34 from FIR Digital Filters as

Equation:

$$C_s(\omega) = \sum_{n=-\infty}^{\infty} x(n/M)\ \text{Ш}_M(n) e^{-j\omega n} = \sum_m x(m) e^{-j\omega Mm}$$

Equation:

$$= C(M\omega)$$

where $C(\omega)$ is the DTFT of $x(n)$. Here again the transforms of the up–sampled signal is the same as the original signal except over M periods. This shows up here as $C_s(\omega)$ being a compressed version of M periods of $C(\omega)$.

The z-transform of an up–sampled sequence is simply derived by:

Equation:

$$Y(z) = \sum_{n=-\infty}^{\infty} y(n) z^{-n} = \sum_n x(n/M) \; \text{Ш}_M(n) z^{-n} = \sum_m x(m) z^{-Mm}$$

Equation:

$$= X\left(z^M\right)$$

which is consistent with a complex version of the DTFT in [link].

Notice that in all of these cases, there is no loss of information or invertibility. In other words, there is no aliasing.

Down–Sampling, Subsampling, or Decimation

In this section we consider the sampling problem where, unless there is sufficient redundancy, there will be a loss of information caused by removing data in the time domain and aliasing in the frequency domain.

The sampling process or the down sampling process creates a new shorter or compressed signal by keeping every M^{th} sample of the original sequence. This process is best seen as done in two steps. The first is to mask off the terms to be removed by setting $M-1$ terms to zero in each length-M block (multiply $x(n)$ by $\text{Ш}_M(n)$), then that sequence is compressed or shortened by removing the $M-1$ zeroed terms.

We will now calculate the length $L = N/M$ DFT of a sequence that was obtained by sampling every M terms of an original length-N sequence $x(n)$. We will use the orthogonal properties of the basis vectors of the DFT which says:
Equation:

$$\sum_{n=0}^{M-1} e^{-j2\pi nl/M} = \begin{cases} M & \text{if n is an integer multiple of} M \\ 0 & \text{otherwise.} \end{cases}$$

We now calculate the DFT of the down-sampled signal.
Equation:

$$C_d\left(k\right) = \sum_{m=0}^{L-1} x\left(Mm\right) W_L^{mk}$$

where $N = LM$ and $k =$ 0,1,...,L-1. This is done by masking $x(n)$.

Equation:

$$C_d\left(k\right) = \sum_{n=0}^{N-1} x\left(n\right) x_M\left(n\right) W_L^{nk}$$

Equation:

$$= \sum_{n=0}^{N-1} x\left(n\right) \left[\frac{1}{M} \sum_{l=0}^{M-1} e^{-j2\pi nl/M}\right] e^{-j2\pi nk/N}$$

Equation:

$$= \frac{1}{M} \sum_{l=0}^{M-1} \sum_{n=0}^{N-1} x\left(n\right) e^{j2\pi(k+Ll)n/N}$$

Equation:

$$= \frac{1}{M} \sum_{l=0}^{M-1} C\left(k + Ll\right)$$

The compression or removal of the masked terms is achieved in the frequency domain by using $k =$ 0,1,...,L-1 This is a length-L DFT of the samples of $x(n)$. Unless $C(k)$ is sufficiently bandlimited, this causes aliasing and $x(n)$ is not unrecoverable.

It is instructive to consider an alternative derivation of the above result. In this case we use the IDFT given by

Equation:

$$x\left(n\right)=\frac{1}{N}\sum_{k=0}^{N-1}C\left(k\right)W_N^{-nk}.$$

The sampled signal gives
Equation:

$$y\left(n\right)=x\left(Mn\right)=\frac{1}{N}\sum_{k=0}^{N-1}C\left(k\right)W_N^{-Mnk}.$$

for $n=0,1,\cdots,L-1$. This sum can be broken down by
Equation:

$$y\left(n\right)=\frac{1}{N}\sum_{k=0}^{L-1}\sum_{l=0}^{M-1}C\left(k+Ll\right)W_N^{-Mn(k+Ll)}.$$

Equation:

$$=\frac{1}{N}\sum_{k=0}^{L-1}\left[\sum_{l=0}^{M-1}C\left(k+Ll\right)\right]W_N^{-Mnk}.$$

From the term in the brackets, we have
Equation:

$$C_s\left(k\right)=\sum_{l=0}^{M-1}C\left(k+Ll\right)$$

as was obtained in [link].

Now consider still another derivation using shah functions. Let
Equation:

$$x_s\left(n\right)=\text{Ш}_M\left(n\right)x\left(n\right)$$

From the convolution property of the DFT we have

Equation:

$$C_s(k) = L\ \text{Ш}_L(k) * C(k)$$

therefore

Equation:

$$C_s(k) = \sum_{l=0}^{M-1} C(k + Ll)$$

which again is the same as in [link].

We now turn to the down sampling of an infinitely long signal which will require use of the DTFT of the signals.

Equation:

$$C_s(\omega) = \sum_{m=-\infty}^{\infty} x(Mm)\, e^{-j\omega Mm}$$

Equation:

$$= \sum_n x(n)\ \text{Ш}_M(n)\, e^{-j\omega n}$$

Equation:

$$= \sum_n x(n) \left[\frac{1}{M} \sum_{l=0}^{M-1} e^{-j2\pi nl/M} \right] e^{-j\omega n}$$

Equation:

$$= \frac{1}{M} \sum_{l=0}^{M-1} \sum_n x(n)\, e^{-j(\omega - 2\pi l/M)n}$$

Equation:

$$= \frac{1}{M} \sum_{l=0}^{M-1} C\left(\omega - 2\pi l/M\right)$$

which shows the aliasing caused by the masking (sampling without compression). We now give the effects of compressing $x_s\left(n\right)$ which is a simple scaling of ω. This is the inverse of the stretching results in [link].

Equation:

$$C_s\left(\omega\right) = \frac{1}{M} \sum_{l=0}^{M-1} C\left(\omega/M - 2\pi l/M\right).$$

In order to see how the various properties of the DFT can be used, consider an alternate derivation which uses the IDTFT.

Equation:

$$x\left(n\right) = \frac{1}{2\pi} \int_{-\pi}^{\pi} C\left(\omega\right) e^{j\omega n} \, d\omega$$

which for the down–sampled signal becomes

Equation:

$$x\left(Mn\right) = \frac{1}{2\pi} \int_{-\pi}^{\pi} C\left(\omega\right) e^{j\omega Mn} \, d\omega$$

The integral broken into the sum of M sections using a change of variables of $\omega = (\omega_1 + 2\pi l)/M$ giving

Equation:

$$x\left(Mn\right) = \frac{1}{2\pi} \sum_{l=0}^{M-1} \int_{-\pi}^{\pi} C\left(\omega_1/M + 2\pi l/M\right) e^{j(\omega_1/M + 2\pi l/M)Mn} \, d\omega_1$$

which shows the transform to be the same as given in Equation 9 from Chebyshev of Equal Ripple Error Approximation Filters.

Still another approach which uses the shah function can be given by
Equation:

$$x_s(n) = \text{Ш}_M(n)\, x(n)$$

which has as a DTFT
Equation:

$$C_s(\omega) = \left(\frac{2\pi}{M}\right)\ \text{Ш}_{2\pi/M}(\omega) * C(\omega)$$

Equation:

$$= \frac{2\pi}{M} \sum_{l=0}^{M-1} C(\omega + 2\pi l/M)$$

which after compressing becomes
Equation:

$$C_s = \frac{2\pi}{M} \sum_{l=0}^{M-1} C(\omega/M + 2\pi l/M)$$

which is same as [Equation 9 from Chebyshev of Equal Ripple Error Approximation Filters](#).

Now we consider the effects of down–sampling on the z-transform of a signal.
Equation:

$$X(z) = \sum_{n=-\infty}^{\infty} x(n)\, z^{-n}$$

Applying this to the sampled signal gives
Equation:

$$X_s(z) = \sum_n x(Mn)\, z^{-Mn} = \sum_n x(n)\ \text{Ш}_M(n)\, z^{-n}$$

Equation:

$$= \sum_{n} x\left(n\right)\ \sum_{l=0}^{M-1} e^{j2\pi nl/M}\ z^{-n}$$

Equation:

$$= \sum_{l=0}^{M-1} \sum_{n} x\left(n\right) \left\{e^{j2\pi l/M}\ z\right\}^{-n}$$

Equation:

$$= \sum_{l=0}^{M-1} X\left(e^{-j2\pi l/M}\ z\right)$$

which becomes after compressing
Equation:

$$= \sum_{l=0}^{M-1} X\left(e^{-j2\pi l/M}\ z^{1/M}\right).$$

This concludes our investigations of the effects of down–sampling a discrete–time signal and we discover much the same aliasing properties as in sampling a continuous–time signal. We also saw some of the mathematical steps used in the development.

More Later

We will later develop relations of sampling to multirate systems, periodically time varying systems, and block processing. This should be a very effective formulation for teaching as well as research on these topics.